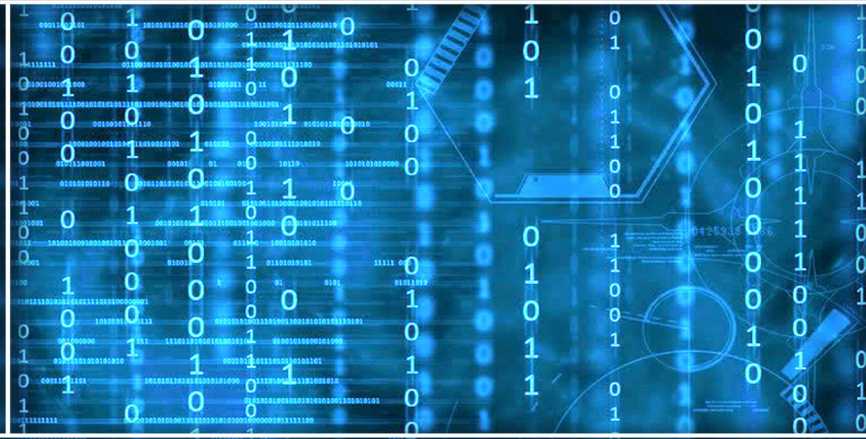


Volume 14 Issue 1

January 2023



ISSN 2156-5570(Online)

ISSN 2158-107X(Print)



# 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 14 Issue 1 January 2023**  
**ISSN 2156-5570 (Online)**  
**ISSN 2158-107X (Print)**

# Editorial Board

## Editor-in-Chief

### **Dr. Kohei Arai - Saga University**

*Domains of Research: Technology Trends, Computer Vision, Decision Making, Information Retrieval, Networking, Simulation*

---

## Associate Editors

### **Alaa Sheta**

#### **Southern Connecticut State University**

*Domain of Research: Artificial Neural Networks, Computer Vision, Image Processing, Neural Networks, Neuro-Fuzzy Systems*

### **Domenico Ciuonzo**

#### **University of Naples, Federico II, Italy**

*Domain of Research: Artificial Intelligence, Communication, Security, Big Data, Cloud Computing, Computer Networks, Internet of Things*

### **Doroła Kaminska**

#### **Lodz University of Technology**

*Domain of Research: Artificial Intelligence, Virtual Reality*

### **Elena Scutelnicu**

#### **"Dunarea de Jos" University of Galati**

*Domain of Research: e-Learning, e-Learning Tools, Simulation*

### **In Soo Lee**

#### **Kyungpook National University**

*Domain of Research: Intelligent Systems, Artificial Neural Networks, Computational Intelligence, Neural Networks, Perception and Learning*

### **Krassen Stefanov**

#### **Professor at Sofia University St. Kliment Ohridski**

*Domain of Research: e-Learning, Agents and Multi-agent Systems, Artificial Intelligence, e-Learning Tools, Educational Systems Design*

### **Renato De Leone**

#### **Università di Camerino**

*Domain of Research: Mathematical Programming, Large-Scale Parallel Optimization, Transportation problems, Classification problems, Linear and Integer Programming*

### **Xiao-Zhi Gao**

#### **University of Eastern Finland**

*Domain of Research: Artificial Intelligence, Genetic Algorithms*

# CONTENTS

Paper 1: Eye-tracking Analysis: College Website Visual Impact on Emotional Responses Reflected on Subconscious Preferences

*Authors: Hedda Martina Šola, Fayyaz Hussain Qureshi, Sarwar Khawaja*

**PAGE 1 – 11**

Paper 2: Improving MapReduce Speculative Executions with Global Snapshots

*Authors: Ebenezer Komla Gavua, Gabor Kecskemeti*

**PAGE 12 – 22**

Paper 3: Recognizing Safe Drinking Water and Predicting Water Quality Index using Machine Learning Framework

*Authors: Mohamed Torky, Ali Bakhiet, Mohamed Bakrey, Ahmed Adel Ismail, Ahmed I. B. EL Seddawy*

**PAGE 23 – 33**

Paper 4: Leaf-based Classification of Important Indigenous Tree Species by Different Feature Extraction Techniques and Selected Classification Algorithms

*Authors: Eugene Val D. Mangaoang, Jaime M. Samaniego*

**PAGE 34 – 41**

Paper 5: A Novel Deep-learning based Approach for Automatic Diacritization of Arabic Poems using Sequence-to-Sequence Model

*Authors: Mohamed S. Mahmoud, Nermin Negied*

**PAGE 42 – 46**

Paper 6: An Investigation of Cybersecurity Issues of Remote Work during the COVID-19 Pandemic in Saudi Arabia

*Authors: Gaseb N Alotibi, Abdulwahid Al Abdulwahid*

**PAGE 47 – 53**

Paper 7: Patent Text Classification based on Deep Learning and Vocabulary Network

*Authors: Ran Li, Wangke Yu, Qianliang Huang, Yuying Liu*

**PAGE 54 – 61**

Paper 8: IoT Technology for Intelligent Management of Energy, Equipment and Security in Smart House

*Authors: Fangmin Yuan, Yan Zhang, Junchao Zhang*

**PAGE 62 – 76**

Paper 9: The Effect of Augmented Reality Mobile Application on Visitor Impact Mediated by Rational Hedonism: Evidence from Subak Museum

*Authors: Ketut Agustini, Dessy Seri Wahyuni, I Nengah Eka Mertayasa, Ni Made Ratminingsih, Gede Ariadi*

**PAGE 77 – 88**

Paper 10: Developing a Computer Simulation to Study the Behavior of Factors Affecting the Flooding of the Gash River

*Authors: Abdalilah G. I. Alhalangy*

**PAGE 89 – 93**

Paper 11: Pinpointing Factors in the Success of Integrated Information System Toward Open Government Data Initiative: A Perspective from Employees

*Authors: Wahyu Setiawan Wibowo, Ahmad Fadhil, Dana Indra Sensuse, Sofian Lusa, Prasetyo Adi Wibowo Putro, Alivia Yulfitri*

**PAGE 94 – 109**

- Paper 12: Autism Spectrum Disorder Detection: Video Games based Facial Expression Diagnosis using Deep Learning  
*Authors: Morched Derbali, Mutasem Jarrah, Princy Randhawa*  
**PAGE 110 – 119**
- Paper 13: An Improved SVM Method for Movement Recognition of Lower Limbs by MIMU and sEMG  
*Authors: Xu Yun, Xu Ling, Gao Lei, Liu Zhanhao, Shen Bohan*  
**PAGE 120 – 124**
- Paper 14: Bidirectional Recurrent Neural Network based on Multi-Kernel Learning Support Vector Machine for Transformer Fault Diagnosis  
*Authors: Xun Zhao, Shuai Chen, Ke Gao, Lin Luo*  
**PAGE 125 – 135**
- Paper 15: DataOps Lifecycle with a Case Study in Healthcare  
*Authors: Shaimaa Bahaa, Atef Z. Ghalwash, Hany Harb*  
**PAGE 136 – 144**
- Paper 16: Evaluation of e-Service Quality Impacts Customer Satisfaction: One-Gate Integrated Service Application in Indonesian Weather Agency  
*Authors: Aji Prasetyo, Deny Irawan, Dana Indra Sensuse, Sofian Lusa, Prasetyo Adi Wibowo, Alivia Yulfitri*  
**PAGE 145 – 152**
- Paper 17: Investigating the Input Validation Vulnerabilities in C Programs  
*Authors: Shouki A. Ebad*  
**PAGE 153 – 160**
- Paper 18: An Optimized Method for Polar Code Construction  
*Authors: Issame El Kaime, Reda Benkhrouya, Abdessalam Ait Madi, Hassane Erguig*  
**PAGE 161 – 165**
- Paper 19: Current Multi-factor of Authentication: Approaches, Requirements, Attacks and Challenges  
*Authors: Ali Hameed Yassir Mohammed, Rudzidatul Akmam Dzyiauddin, Liza Abdul Latiff*  
**PAGE 166 – 178**
- Paper 20: Arabic Dialogue Processing and Act Classification using Support Vector Machine  
*Authors: Abraheem Mohammed Sulayman Alsubayhay, Md Sah Hj Salam, Farhan Bin Mohamed*  
**PAGE 179 – 190**
- Paper 21: Tree-based Machine Learning and Deep Learning in Predicting Investor Intention to Public Private Partnership  
*Authors: Ahmad Amin, Rahmawaty, Maya Febrianty Lautania, Suraya Masrom, Rahayu Abdul Rahman*  
**PAGE 191 – 195**
- Paper 22: VHDL based Design of an Efficient Hearing Aid Filter using an Intelligent Variable-Bandwidth-Filter  
*Authors: Ujjwala S Rawandale, Sanjay R. Ganorkar, Mahesh T. Kolte*  
**PAGE 196 – 200**
- Paper 23: Performance Evaluation of Photovoltaic Projects in Latin America  
*Authors: Cristian León-Ospina, Heyner Arias-Zarate, Cesar Hernandez*  
**PAGE 201 – 212**

**Paper 24: Ransomware: Analysis of Encrypted Files**

*Authors: Houria MADANI, Noura OUERDI, Abdelmalek Azizi*

**PAGE 213 – 217**

**Paper 25: MMZ: A Study on the Implementation of Mathematical Game-based Learning Tool**

*Authors: Nur Syaheera Binti Sulaiman, Hamzah Asyrani Bin Sulaiman, Nor Saradatul Akmar Binti Zulkifli, Tuty Asmawanty Binti Abdul Kadir*

**PAGE 218 – 224**

**Paper 26: An Improved Poisson Surface Reconstruction Algorithm based on the Boundary Constraints**

*Authors: Zhouqi Liu, Lei Wang, Muhammad Tahir, Jin Huang, Tianqi Cheng, Xiping Guo, Yuwei Wang, ChunXiang Liu*

**PAGE 225 – 232**

**Paper 27: Data Augmentation for Deep Learning Algorithms that Perform Driver Drowsiness Detection**

*Authors: Ghulam Masudh Mohamed, Sulaiman Saleem Patel, Nalindren Naicker*

**PAGE 233 – 248**

**Paper 28: An Improved Ant Colony Algorithm for Virtual Resource Scheduling in Cloud Computing**

*Authors: Chunlei Zhong, Gang Yang*

**PAGE 249 – 261**

**Paper 29: Deep Analysis of Risks and Recent Trends Towards Network Intrusion Detection System**

*Authors: D. Shankar, G. Victo Sudha George, Janardhana Naidu J N S S, P Shyamala Madhuri*

**PAGE 262 – 276**

**Paper 30: Qdijbo Metaheuristic Optimization Algorithm for Computation of Real-Parameters and Engineering Design Optimization**

*Authors: Ikpo C V, Akowuah E K, Kponyo J J, Boateng K O*

**PAGE 277 – 297**

**Paper 31: Modeling of Whale Optimization with Deep Learning based Brain Disorder Detection and Classification**

*Authors: Uvaneshwari M, M. Baskar*

**PAGE 298 – 305**

**Paper 32: A Study on the Designation Institution for Supercomputer Specialized Centers in Republic of Korea**

*Authors: Hyungwook Shim, Yonghwan Jung, Jaegyeon Hahm*

**PAGE 306 – 312**

**Paper 33: Time Series Forecasting using LSTM and ARIMA**

*Authors: Khulood Albeladi, Bassam Zafar, Ahmed Mueen*

**PAGE 313 – 320**

**Paper 34: Enhancing the Intrusion Detection Efficiency using a Partitioning-based Recursive Feature Elimination in Big Cloud Environment**

*Authors: Hesham M. Elmasry, Ayman E. Khedr, Hatem M. Abdelkader*

**PAGE 321 – 330**

**Paper 35: Parameter Extraction and Performance Analysis of 3D Surface Reconstruction Techniques**

*Authors: Richha Sharma, Pawanesh Abrol*

**PAGE 331 – 336**

**Paper 36: Classifying Weather Images using Deep Neural Networks for Large Scale Datasets**

*Authors: Shweta Mittal, Om Prakash Sangwan*

**PAGE 337 – 343**

**Paper 37: Exploring College Academic Performance of K to12 IT and Non-IT-related Strands to Reduce Academic Deficiencies**

*Authors: Marilou S. Benozza, Thelma Palaoag*

**PAGE 344 – 355**

**Paper 38: AI based Dynamic Prediction Model for Mobile Health Application System**

*Authors: Adari Ramesh, C K Subbaraya, G K Ravi Kumar*

**PAGE 356 – 365**

**Paper 39: Reducing Cheating in Online Exams Through the Proctor Test Model**

*Authors: Yusring Sanusi Baso, Nurul Murtadho, Syihabuddin, Hikmah Maulani, Andi Agussalim, Haeruddin, Ahmad Fadlan, Ilham Ramadhan*

**PAGE 366 – 370**

**Paper 40: A Framework of Outcome-based Assessment and Evaluation for Computing Programs**

*Authors: Wasan S. Awad, Khadija A. Almhoson*

**PAGE 371 – 380**

**Paper 41: A Hybrid Filtering Technique of Digital Images in Multimedia Data Warehouses**

*Authors: Nermin Abdelhakim Othman, Ahmed Ayman Saad, Ahmed Sharaf Eldin*

**PAGE 381 – 392**

**Paper 42: Agro-Food Supply Chain Traceability using Blockchain and IPFS**

*Authors: Subashini Babu, Hemavathi Devarajan*

**PAGE 393 – 399**

**Paper 43: Implementation of Flood Emergency Response System with Face Analytics**

*Authors: E. Mardaid, Z. Zainal Abidin, S. A. Asmai, Z. Abal Abas*

**PAGE 400 – 406**

**Paper 44: A Novel Smart Deepfake Video Detection System**

*Authors: Marwa Elpeltagy, Aya Ismail, Mervat S. Zaki, Kamal Eldahshan*

**PAGE 407 – 419**

**Paper 45: Three on Three Optimizer: A New Metaheuristic with Three Guided Searches and Three Random Searches**

*Authors: Purba Daru Kusuma, Ashri Dinimaharawati*

**PAGE 420 – 429**

**Paper 46: An Empirical Study on the Affecting Factors of Cloud-based ERP System Adoption in Iraqi SMEs**

*Authors: Mohammed G. J, MA Burhanuddin, Dawood F. A. A, Alyousif S, Alkhayyat A, Ali M. H, R. Q. Malik, Jaber M. M*

**PAGE 430 – 441**

**Paper 47: A Learning-based Correlated Graph Model for Spinal Cord Injury Prediction from Magnetic Resonance Spinal Images**

*Authors: P. R. S. S. V Raju, V. Asanambigai, Suresh Babu Mudunuri*

**PAGE 442 – 449**

Paper 48: The Practices of Online Assessment in a Digital Device in the Context of University Training: The Case of Hassan II University

*Authors: Fatima-ezzahra Mrisse, Nadia Chafiq, Mohammed Talbi, Kamal Moundy*

**PAGE 450 – 455**

Paper 49: Comparative and Evaluation of Anomaly Recognition by Employing Statistic Techniques on Humanoid Robot

*Authors: Nuratiqa Natrah Mansor, Muhammad Herman Jamaluddin, Ahmad Zaki Shukor, Muhammad Sufyan Basri*

**PAGE 456 – 464**

Paper 50: Queueing Model based Dynamic Scalability for Containerized Cloud

*Authors: Ankita Srivastava, Narander Kumar*

**PAGE 465 – 472**

Paper 51: An Improved Breast Cancer Classification Method Using an Enhanced AdaBoost Classifier

*Authors: Yousef K. Qawqzeh, Abdullah Alourani, Sameh Ghwanmeh*

**PAGE 473 – 478**

Paper 52: Efficient Multimedia Content Transmission Model for Disaster Management using Delay Tolerant Mobile Adhoc Networks

*Authors: Sushant Mangasuli, Mahesh Kaluti*

**PAGE 479 – 484**

Paper 53: The Effect of Thermal and Electrical Conductivities on the Ablation Volume during Radiofrequency Ablation Process

*Authors: Mohammed S. Ahmed, Mohamed Tarek El-Wakad, Mohammed A. Hassan*

**PAGE 485 – 493**

Paper 54: A Light-weight Authentication Scheme in the Internet of Things using the Enhanced Bloom Filter

*Authors: Xiaoyan Huo*

**PAGE 494 – 501**

Paper 55: Generalized Epileptic Seizure Prediction using Machine Learning Method

*Authors: Zarqa Altaf, Mukhtiar Ali Unar, Sanam Narejo, Muhammad Ahmed Zaki, Naseer-u-Din*

**PAGE 502 – 510**

Paper 56: The Impact of COVID-19 on Digital Competence

*Authors: Syerina Syahrin, Khalid Almashiki, Eman Alzaanin*

**PAGE 511 – 519**

Paper 57: A Survey on Cloudlet Computation Optimization in the Mobile Edge Computing Environment

*Authors: Layth Muwafaq, Nor K. Noordin, Mohamed Othman, Alyani Ismail, Fazirulhisyam Hashim*

**PAGE 520 – 532**

Paper 58: Proof-of-Work for Merkle based Access Tree in Patient Centric Data

*Authors: B Ravinder Reddy, T Adilakshmi*

**PAGE 533 – 539**

Paper 59: Water Tank Wudhu and Monitoring System Design using Arduino and Telegram

*Authors: Ritkhal, Yuggo Afrianto, Indra Riawan, Fitrah Satrya Fajar Kusumah, Dwi Remawati*

**PAGE 540 – 546**



Paper 60: Risk Analysis of Urban Water Infrastructure Systems in Cauayan City

Authors: Rafael J. Padre, Melanie A. Baguio, Edward B. Panganiban, Rudy U. Panganiban, Carluz R. Bautista, Justine Ryan L. Rigates, Allisandra Pauline Mariano

PAGE 547 – 558

Paper 61: Mitigate Volumetric DDoS Attack using Machine Learning Algorithm in SDN based IoT Network Environment

Authors: Kumar J, Arul Leena Rose P J

PAGE 559 – 568

Paper 62: Customer Sentiment Analysis in Hotel Reviews Through Natural Language Processing Techniques

Authors: Soumaya Ounacer, Driss Mhamdi, Soufiane Ardchir, Abderrahmane Daif, Mohamed Azzouazi

PAGE 569 – 579

Paper 63: Performance Comparison of the Kernels of Support Vector Machine Algorithm for Diabetes Mellitus Classification

Authors: Dimas Aryo Anggoro, Dian Permatasari

PAGE 580 – 585

Paper 64: Image Segmentation of Intestinal Polyps using Attention Mechanism based on Convolutional Neural Network

Authors: Xinyi Zheng, Wanru Gong, Ruijia Yang, Guoyu Zuo

PAGE 586 – 593

Paper 65: A Novel Hybrid DL Model for Printed Arabic Word Recognition based on GAN

Authors: Yazan M. Alwaqfi, Mumtazimah Mohamad, Ahmad T. Al-Taani, Nazirah Abd Hamid

PAGE 594 – 604

Paper 66: Quantum Cryptography Experiment using Optical Devices

Authors: Nur Shahirah Binti Azahari, Nur Ziadah Binti Harun

PAGE 605 – 610

Paper 67: Analysis of Medical Slide Images Processing using Depth Learning in Histopathological Studies of Cerebellar Cortex Tissue

Authors: Xiang-yu Zhang, Xiao-wen Shi, Xing-bo Zhang

PAGE 611 – 621

Paper 68: The Cloud-powered Hybrid Learning Process to Enhance Digital Natives' Analytical Reading Skills

Authors: Sakolwan Napaporn, Sorakrich Maneewan, Kuntida Thamwipat, Vitsanu Nittayathamkul

PAGE 622 – 627

Paper 69: A Model for Detecting Fungal Diseases in Cotton Cultivation using Segmentation and Machine Learning Approaches

Authors: Odukoya O. H, Aina S, Dégbéssé F. W

PAGE 628 – 636

Paper 70: Deep Learning Models for the Detection of Monkeypox Skin Lesion on Digital Skin Images

Authors: Othman A. Alrusaini

PAGE 637 – 644

Paper 71: Upgraded Very Fast Decision Tree: Energy Conservative Algorithm for Data Stream Classification

Authors: Mai Lefa, Hatem Abd-Elkader, Rashed Salem

PAGE 645 – 652

**Paper 72: Classification Model for Diabetes Mellitus Diagnosis based on K-Means Clustering Algorithm Optimized with Bat Algorithm**

*Authors: Syaiful Anam, Zuraidah Fitriah, Noor Hidayat, Mochamad Hakim Akbar Assidiq Maulana*

**PAGE 653 – 659**

**Paper 73: Descriptive Analytics and Interactive Visualizations for Performance Monitoring of Extension Services Programs, Projects, and Activities**

*Authors: Noelyn M. De Jesus, Lorissa Joana E. Buenas*

**PAGE 660 – 668**

**Paper 74: Arabic Stock-News Sentiments and Economic Aspects using BERT Model**

*Authors: Eman Alasmari, Mohamed Hamdy, Khaled H. Alyoubi, Fahd Saleh Alotaibi*

**PAGE 669 – 679**

**Paper 75: e-Government Usability Evaluation: A Comparison between Algeria and the UK**

*Authors: Mohamed Benaida*

**PAGE 680 – 690**

**Paper 76: The Effect of Artificial Neural Network Towards the Number of Particles of Rao-Blackwellized Particle Filter using Laser Distance Sensor**

*Authors: Amirul Jamaludin, Norhidayah Mohamad Yatim, Zarina Mohd Noh*

**PAGE 691 – 700**

**Paper 77: Expanding Louvain Algorithm for Clustering Relationship Formation**

*Authors: Murniyati, Achmad Benny Mutiara, Setia Wirawan, Tristyanti Yusnitasari, Dyah Anggraini*

**PAGE 701 – 708**

**Paper 78: Implementation Failure Recovery Mechanism using VLAN ID in Software Defined Networks**

*Authors: Heru Nurwarsito, Galih Prasetyo*

**PAGE 709 – 714**

**Paper 79: Analysis of the Artificial Neural Network Approach in the Extreme Learning Machine Method for Mining Sales Forecasting Development**

*Authors: Hendra Kurniawan, Joko Triloka, Yunus Ardhan*

**PAGE 715 – 721**

**Paper 80: Deca Convolutional Layer Neural Network (DCL-NN) Method for Categorizing Concrete Cracks in Heritage Building**

*Authors: Dinar Mutiara Kusumo Nugraheni, Andi Kurniawan Nugroho, Diah Intan Kusumo Dewi, Beta Noranita*

**PAGE 722 – 730**

**Paper 81: Convolutional Transformer based Local and Global Feature Learning for Speech Enhancement**

*Authors: Chaitanya Jannu, Sunny Dayal Vanambathina*

**PAGE 731 – 743**

**Paper 82: Assessing User Interest in Web API Recommendation using Deep Learning Probabilistic Matrix Factorization**

*Authors: T. Ramathulasi, M. Rajasekhara Babu*

**PAGE 744 – 752**

**Paper 83: Unsupervised Learning-based New Seed-Expanding Approach using Influential Nodes for Community Detection in Social Networks**

*Authors: Khaoula AIT RAI, Mustapha MACHKOUR, Jilali ANTARI*

**PAGE 753 – 766**

**Paper 84: Implementation of ICT Continuity Plan (ICTCP) in the Higher Education Institutions (HEI'S): SUC'S Awareness and its Status**

*Authors: Chester L. Cofino, Ken M. Balogo, Jeffrey G. Alegia, Michael Marvin P. Cruz, Benjamin B. Alejado Jr, Felicisimo V. Wenceslao Jr*

**PAGE 767 – 771**

**Paper 85: Towards a Machine Learning-based Model for Automated Crop Type Mapping**

*Authors: Asmae DAKIR, Fatimazahra BARRAMOU, Omar Bachir ALAMI*

**PAGE 772 – 779**

**Paper 86: A Hybrid Model by Combining Discrete Cosine Transform and Deep Learning for Children Fingerprint Identification**

*Authors: Vaishali Kamble, Manisha Dale, Vinayak Bairagi*

**PAGE 780 – 787**

**Paper 87: 2-D Deep Convolutional Neural Network for Predicting the Intensity of Seismic Events**

*Authors: Assem Turarbek, Yeldos Adetbekov, Maktagali Bektemesov*

**PAGE 788 – 796**

**Paper 88: User-Centered Design (UCD) of Time-Critical Weather Alert Application**

*Authors: Abdulelah M. Ali, Abdulrahman Khamaj, Ziho Kang, Majed Moosa, Mohd Mukhtar Alam*

**PAGE 797 – 808**

**Paper 89: Interventional Teleoperation Protocol that Considers Stair Climbing or Descending of Crawler Robots in Low Bit-rate Communication**

*Authors: Tsubasa Sakaki, Kei Sawai*

**PAGE 809 – 817**

**Paper 90: Business Intelligence Data Visualization for Diabetes Health Prediction**

*Authors: Samantha Siow Jia Qi, Sarasvathi Nagalingham*

**PAGE 818 – 831**

**Paper 91: Augmented, Virtual and Mixed Reality Research in Cultural Heritage: A Bibliometric Study**

*Authors: Nilam Upasani, Asmita Manna, Manjiri Ranjanikar*

**PAGE 832 – 842**

**Paper 92: Implementation of Business Intelligence Solution for United Airlines**

*Authors: Ng Iris, Sarasvathi Nagalingham*

**PAGE 843 – 852**

**Paper 93: Model Predictive Controlled Quasi Z Source Inverter Fed Induction Motor Drive System**

*Authors: D. Himabindu, G. Sreenivasan, R. Kiranmayi*

**PAGE 853 – 856**

**Paper 94: Visualization of Business Intelligence Insights into Aviation Accidents**

*Authors: Loe Piin Piin, Sarasvathi Nagalingham*

**PAGE 857 – 875**

**Paper 95: Metaphor Recognition Method based on Graph Neural Network**

*Authors: Zhou Chuwei, SHI Yunmei*

**PAGE 876 – 883**

**Paper 96: User Perceive Realism of Machine Learning-based Drone Dynamic Simulator**

*Authors: Damitha Sandaruwan, Nihal Kodikara, Piyumi Radeeshani, K.T.Y. Mahima, Chathura Suduwella, Sachintha Pitigala, Mangalika Jayasundara*

**PAGE 884 – 893**

**Paper 97: Stacking Deep-Learning Model, Stories and Drawing Properties for Automatic Scene Generation**

*Authors: Samir Elloumi, Nzamba Bignoumba*

**PAGE 894 – 911**

**Paper 98: A Machine Learning Hybrid Approach for Diagnosing Plants Bacterial and Fungal Diseases**

*Authors: Ahmed BaniMustafa, Hazem Qattous, Ihab Ghabeish, Muwaffaq Karajeh*

**PAGE 912 – 921**

**Paper 99: Enhancing Collaborative Interaction with the Augmentation of Sign Language for the Vocally Challenged**

*Authors: Sukruth G L, Vijaya Kumar B P, Tejas M R, Rithvik K, Trisha Ann Tharakan*

**PAGE 922 – 929**

**Paper 100: Delivery Management System based on Blockchain, Smart Contracts and NFT: A Case Study in Vietnam**

*Authors: Khiem Huynh Gia, Luong Hoang Huong, Hong Khanh Vo, Phuc Nguyen Trong, Khoa Tran Dang, Hieu Le Van, Loc Van Cao Phu, Duy Nguyen Truong Quoc, Nguyen Huyen Tran, Anh Nguyen The, Huynh Trong Nghia, Bang Le Khanh, Kiet Le Tuan, Nguyen Thi Kim Ngan*

**PAGE 930 – 938**

**Paper 101: Integrated Assessment of Teaching Efficacy: A Natural Language Processing Approach**

*Authors: Lalitha Manasa Chandrapati, Ch. Koteswara Rao*

**PAGE 939 – 947**

**Paper 102: An Effect Assessment System for Curriculum Ideology and Politics based on Students' Achievements in Chinese Engineering Education**

*Authors: Bo Wang, Hailuo Yu, Yusheng Sun, Zhifeng Zhang, Xiaoyun Qin*

**PAGE 948 – 953**

**Paper 103: Navigation of Autonomous Vehicles using Reinforcement Learning with Generalized Advantage Estimation**

*Authors: Edwar Jacinto, Fernando Martinez, Fredy Martinez*

**PAGE 954 – 959**

**Paper 104: A Low-Cost Wearable Autonomous System for the Protection of Bicycle Users**

*Authors: Daniel Mejia, Sergio Gomez, Fredy Martinez*

**PAGE 960 – 966**

**Paper 105: An Automated Impact Analysis Approach for Test Cases based on Changes of Use Case based Requirement Specifications**

*Authors: Adisak Intana, Kanjana Laosen, Thiwatip Sriraksa*

**PAGE 967 – 980**

**Paper 106: Trust Management for Deep Autoencoder based Anomaly Detection in Social IoT**

*Authors: Rashmi M R, C Vidya Raj*

**PAGE 981 – 989**

**Paper 107: Machine Learning Techniques to Enhance the Mental Age of Down Syndrome Individuals: A Detailed Review**

*Authors: Irfan M. Leghari, Hamimah Ujir, SA Ali, Irwandi Hipiny*

**PAGE 990 – 999**

**Paper 108: AMIM: An Adaptive Weighted Multimodal Integration Model for Alzheimer's Disease Classification**

*Authors: Dewen Ding, Xianhua Zeng, Xinyu Wang, Jian Zhang*

**PAGE 1000 – 1007**

# Eye-tracking Analysis: College Website Visual Impact on Emotional Responses Reflected on Subconscious Preferences

Dr Hedda Martina Šola<sup>1</sup>, Dr Fayyaz Hussain Qureshi<sup>2</sup>, Sarwar Khawaja<sup>3</sup>

Centre for Applied Research and Entrepreneurship, Oxford Business College, Oxford, UK<sup>1,2</sup>  
Oxford EducationGroup, Oxford Business College, Oxford, UK<sup>3</sup>

**Abstract**—This study examined students' behaviour on the college website and the content of information they were able to obtain. With the eye-tracking sensor, this study aims to investigate the university websites' effectiveness, satisfaction, and efficiency and collect data regarding users' visual impacts. The research was carried out using mobile phone neuromarketing tools of eye-tracking, facial coding, and supplementary short memory post-survey. The study was focused on two web pages, the homepage, and the CARE page. The analysis results from both web pages were then compared and further discussed. The results suggest that participants mostly elicited sadness (29.55%), neutrality (33.19%), and puzzlement (13.60%) while browsing the homepage, regardless of the areas of interest (AOI). They also elicited slight disgust (4.33%), fear (3.51%), joy (5.21%), and surprise (29.55%). The heat map for the CARE page reveals that the top of the CARE page was a point of attraction for participants. The study found that participants' negative feelings were more intense than good ones concerning homepage scrolling. Also, their pleasant mood intensity increased moderately when they looked at regions with only photos in a subdued color scheme or where brighter colors were used to emphasize essential textual information such as upcoming events and student blogs. This reveals that the website's complexity further affects the cognitive load. Therefore, making it more accessible will be beneficial to students. According to the student's responses, change such as the page's design, color, and text could be implemented.

**Keywords**—Neuromarketing; eye-tracking; student behavior; college website analyses; mood intensities; visual impact; website conversions

## Abbreviation

HCI: Human-Computer Interaction

OSH: Workplace Safety and Health

OBC: Oxford Business College

CARE: Centre for Applied Research and Entrepreneurship

AOI: Areas of Interest

TTF: Time to First Fixation

## I. INTRODUCTION

A college's website is becoming an effective tool for students to collect information in the decision-making process for higher education. The first impressions of potential students are affected by the website. Thus, the digital presence of a

college is crucial since it has been discovered that browsing the website first is a prelude to visiting the campus. Colleges and universities use websites to attract prospective students [1]. Every year, most students decide to pick the higher education institute that will have far-reaching consequences for the rest of their lives. This choice will impact their career, wages, and professional growth. In 2005, private schools spent approximately \$2,073 per new student to apprentice, making communication and recruitment efficiency significant for institutions [2]. Where do awaited students go to learn about colleges and universities? They look through the websites developed by universities and colleges [2]. An appropriate Web-User communication system must be present to catch students' attention, thus, beginning with highlighting the specific users that extensively rely on web pages, the concerns on which the Web page focuses, or the message it wishes to convey to the users [3]. Therefore, it is crucial to investigate the whole process, for instance, how students approach a Web page and how they conclude it. This process must strive for an admiring emotional influence in visual content focused entirely on the student's experience. The basic scheme favors consumers' methods to develop an interactive web design. Eye tracking is an existing tool that allows for carrying out this task. Although its primary purpose has been the retail industry [4,5,6], it has applications in research and commercial grounds underlying human thinking and behavior. Eye tracking is heavily used in studying the influence of imaging to drive new consumers [7,8], but it also has wide use in clinical studies [9,10]. Understanding the visual impact from an emotional perspective is an attractive topic [11] and crucial to different disciplines. However, limited studies attempted to understand students' behavior in learning [12,13] and even lesser to website visual interpretation [14].

Since Students rely on visiting colleges' websites to help make their selection, the aspects influencing their decision are fundamental. Therefore, the current study analyzes the emotional responses to website conversion stimulated by visual impacts. Our main aim is to examine the effectiveness, satisfaction, and efficiency of university websites using eye-tracking analysis and collect data regarding the visual impacts of users. Eye-tracking recognizes the human pupil and traces and analyses the eye movement and fixation when viewing images or websites to allow the investigator to identify where a person is looking at a given period and the arrangement in which their eyes travel from one side to the following [15]. An

eye tracker is a device that measures the location of the eyes [16]. The software will then generate a so-called "heat map" through which the colors will identify the viewers' focused attention [15], permitting investigation of students' emotional reflections on websites.

Due to the development of sophisticated and economical technologies such as wearable sensor techniques, affective computing researchers are increasingly interested in emotion detection. Hence the current study emphasizes the following:

How visual impact can evoke emotional responses reflects website conversion and the eye-tracking analysis of college websites.

This research comprises several steps that initiated the creation of a public link on the OBC official website and served as an invitation to participate. The study was performed using online neuromarketing research technologies from mobile phones for their numerous uses by the students and everybody who visits their website. The experiments were performed in real-time operating neuro metrics combined with supplementary short memory post-survey questions. No additional hardware or application was required due to the advanced development of mobile cameras and machine learning algorithms. Tobii Sticky provided a way to acquire large-scale eye-tracking data.

## II. MATERIALS AND METHODS

A sample of 529 OBC students (both genders, 18-50) participated in the study, exploring how visual impact can evoke emotional responses and reflects on website conversions. Three hundred thirteen participants completed the entire experiment, from which we got one hundred eighty-six usable recordings whose gaze and/or emotion was trackable during their session (i.e., had proper lighting and did not move). One hundred twenty-four participants partially were involved in the study (participants who started the experiment but closed the browser or had timed out before reaching the end), zero participants were excluded from the study, and ninety two were screened out (participants that ended their session based on a screen out the question or did not meet technical requirements).

This study lasted ten days and was accessed via a pop-up banner on the OBC's website specifically created for study purposes. To ensure that the desired level of power and significant results are achieved, the required sample size was calculated a priori using G\*Power [17]. Based on the G\*Power output to detect the effect with 95% power, a two-sided significance level of 5% and a sample size of  $n=23$  for the OBC homepage and  $n=35$  for the CARE page are required (see Appendix). Tested subjects were recruited from the database of Oxford Business College. Participation in the study was voluntary, and no incentives were given.

The Institute for Neuromarketing Ethics Committee approved this research and supervised the study to be underlined with local and international ethical guidelines officially posted on the official Institute's website. Following the British Educational Research Association's ethical guidelines for educational research [18], all participants were informed about the study. They gave their written, informed

consent in digital form before participating in the study. Participants' data were treated according to standard practice and in compliance with GDPR (General Data Protection Regulation).

The online platform for advanced quantitative research, 'Tobii Sticky,' was used to measure eye-tracking and facial coding. According to the results, Sticky's average gaze error in a real-world (non-lab) environment is 1.6 to 1.8 degrees (~5% of the screen width and 7% of screen height) on a laptop which is more than accurate to the vital outcome. Therefore, since this research is conducted outside the laboratory without physical control over the ISO eye-tracking standards, to avoid data obtained from participants who did not meet entirely technical requirements, only recordings labeled as "usable" in Tobii Sticky were utilized. Participants were also provided with a set of images with instructions to ensure compliance with the technical requirements of the eye calibration test. The task started with a 5-point-eye-tracker calibration, a standard procedure for eye-tracking devices. After the calibration, all participants looked for 25 seconds at the identical two website pages (main page and CARE page) at the domain <https://oxfordbusinesscollege.ac.uk/> pre-recorded in the form of the scrollable image, presented in the same order as on the OBC's website page. They could not skip a certain part of the research, but with mouse clicks, it was measured how many participants wanted to stop the research. Such data indicated the total number of participants interested in the content published on the official website of the OBC. While browsing through the two website pages as they usually do, participants' gazing activity and emotional expression were measured over time using webcam-based eye-tracking. After each website page was presented, participants were given five statements to rate, assessing their conscious preferences and opinions about the seen website. They had 10 seconds to do that. The total duration of the study was 1 minute and 93 seconds per link (See 3.1.3.).

Participants' data were treated according to standard practice and in compliance with GDPR and the European Code of Ethics for Research. Since the study was conducted outside a controlled laboratory environment, participants received participation details with the html link. Participants received two 'html' links to participating in this study. One link led to the homepage of the OBC website, while the other led to a subpage related to the brand CARE. They were asked to exclusively access the study from their mobile devices with a working webcam due to the website analytics on how most visits are applied from the mobile phone. We have limited participants to access from mobile; according to the stats, the majority of OBC students participate in the website through mobile devices only.

In the first task, participants were asked to navigate through the OBC's homepage as they usually would, while in the second task, they were asked to browse through the CARE page to find more information about the research resources offered since the goal of the research was threefold: 1) to determine how much the content on the homepage is following the preferences of visitors; 2) how the website can be improved based on the obtained neurometrics; 3) to get an insight into participants' attentiveness levels leading to encoding and recall

of information based on the page design. During this website testing, participants' gazing activity was measured over time to calculate their attention distribution. Heat maps were utilized to observe the movement pattern around the website, and emotions were analyzed through facial coding to obtain insights into participants' behaviour and reactions to the website. In the end, participants were also asked to answer five questions to gauge their user experience, learn about their expectations and reasons for visiting the website, and test their attentiveness levels.

### III. RESULTS

The experiment data obtained from participants were grouped into two subgroups: homepage and CARE page, each representing one website at the official OBC website link: [www.oxfordbusinesscollege.ac.uk](http://www.oxfordbusinesscollege.ac.uk).

#### A. Homepage

The participants' results from the homepage test were investigated through the analysis of the eye tracking data, the facial coding, and the survey data.

1) *Eye-tracking data:* Regarding the eye-tracking measurements, a one-way repeated-measures ANOVA was performed for time to first fixation (TTFF), percentage of participants who fixated on the AOI, number of fixations within AOI, and time spent fixating on the AOI. The heat map for the OBC's homepage reveals that the top of the homepage was a point of attraction for participants (Fig. 1). This is not surprising as the top part is immediately visible once you open the webpage, while to see other information, one needs to scroll the website. The seen order indicates that participants first notice the part of the homepage in the middle of their screen and their eye level height (AOI 2), followed by the upper part of the homepage (AOI 1).

Based on the obtained heat map, in this browsing-only condition, gaze points were concentrated from left to center, meaning that most information presented on the right was not seen. For the quantitative eye-tracking analysis, the OBC's homepage was divided into five areas of interest (AOIs) (Fig. 2).

A one-way within-groups analysis of variance (ANOVA) was used to assess the impact of AOIs of the OBC's homepage page on the attention metrics (i.e., Number of fixations, time to the first fixation, time viewed, Number of visits, Fixating Percentage) (Table I, Table II).

Null hypothesis: There are no significant differences between the number of fixations, number of visits, time to the first fixation, time viewed, and a fixating percentage between AOIs of the homepage.

The result of the analysis showed there was a significant difference within the number of fixations between the AOIs of the OBC's homepage ( $F(5, 180) = 7.223, p = .000$ ). The post hoc analysis demonstrated that the total number of fixations of AOI 1 was significantly higher than that of AOI 3 (mean difference = 21.07), AOI 4 (mean difference = 30.70), and AOI 5 (mean difference = 29.83). The total number of fixations of

AOI 2 was significantly higher than AOI 4 (mean difference = 28.30) and AOI 5 (mean difference = 27.43). The total number of fixations of AOI 6 was not significantly higher or lower than any other AOIs on the homepage.



Fig. 1. Heat map and seen order for the OBC's homepage.

TABLE I. SUMMARISED STATISTICS FOR OBC'S HOMEPAGE

Number of Fixations		$F(5, 180) = 7.223, p = .000$	
AOIs		Mean Difference	p-value
AOI 1	AOI 2	2.40	.999
	AOI 3	21.07	.019*
	AOI 4	30.70	.000*
	AOI 5	29.83	.000*
	AOI 6	18.50	.058



AOI 2	AOI 3	18.67	.054
	AOI 4	28.30	.000*
	AOI 5	27.43	.001*
	AOI 6	16.10	.142
AOI 3	AOI 4	9.63	.682
	AOI 5	8.77	.762
	AOI 6	-2.57	.999
AOI 4	AOI 5	-.87	1.000
	AOI 6	-12.20	.428
AOI 5	AOI 6	-11.33	.513
<i>Number of Visits</i>		F (5, 180) = 9.082, p = .000	
AOI 1	AOI 2	-1.30	.705
	AOI 3	1.93	.274
	AOI 4	3.13	.009*
	AOI 5	2.80	.028*
	AOI 6	3.77	.001*
AOI 2	AOI 3	3.23	.006*
	AOI 4	4.43	.000*
	AOI 5	4.10	.000*
	AOI 6	5.07	.000*
AOI 3	AOI 4	1.20	.771
	AOI 5	.87	.931
	AOI 6	1.83	.333
AOI 4	AOI 5	-.33	.999
	AOI 6	.63	.982
AOI 5	AOI 6	.97	.894
<i>Time to First Fixation</i>		F (5, 180) = 16.205, p = .000	
AOI 1	AOI 2	1.26	.847
	AOI 3	-4.18	.002*
	AOI 4	-7.64	.000*
	AOI 5	-8.04	.000*
	AOI 6	-7.69	.000*
AOI 2	AOI 3	-5.44	.000*
	AOI 4	-8.90	.000*
	AOI 5	-9.30	.000*
	AOI 6	-8.95	.000*
AOI 3	AOI 4	-3.46	.030*
	AOI 5	-3.86	.008*
	AOI 6	-3.51	.058
AOI 4	AOI 5	-.40	.999
	AOI 6	-.05	1.000
AOI 5	AOI 6	.35	1.000
<i>Time Viewed</i>		F (5, 180) = 7.244, p = .000	
AOI 1	AOI 2	.09163	1.000
	AOI 3	1.38	.022*
	AOI 4	2.01	.000*
	AOI 5	1.95	.000*
	AOI 6	1.22	.059
AOI 2	AOI 3	1.29	.040*

	AOI 4	1.92	.000*
	AOI 5	1.86	.000*
	AOI 6	1.13	.101
AOI 3	AOI 4	.63	.693
	AOI 5	.58	.768
	AOI 6	-.15	.999
AOI 4	AOI 5	-.05	1.000
	AOI 6	-.79	.462
AOI 5	AOI 6	-.73	.544
<i>Fixating Percentage</i>		F (5, 180) = 0.896, p = .519	

TABLE II. AVERAGE VALUES OF EYE-TRACKING METRICS FOR OBC'S HOMEPAGE

<i>Time to First Fixation (s)</i>	
<i>AOI</i>	<i>M</i>
AOI 1	1.66
AOI 2	0.40
AOI 3	5.84
AOI 4	9.30
AOI 5	9.70
AOI 6	9.35
<i>Number of Fixations</i>	
AOI 1	1301
AOI 2	1229
AOI 3	669
AOI 4	380
AOI 5	406
AOI 6	746
<i>Number of Visits</i>	
AOI 1	188
AOI 2	227
AOI 3	130
AOI 4	94
AOI 5	104
AOI 6	75
<i>Time Viewed (s)</i>	
AOI 1	2.95
AOI 2	2.76
AOI 3	1.52
AOI 4	1.01
AOI 5	1.00
AOI 6	2.71
<i>Fixating percentage (%)</i>	
AOI 1	96.67
AOI 2	100
AOI 3	96.67
AOI 4	83.33
AOI 5	90
AOI 6	60



Fig. 2. AOIs for OBC's homepage.

A significant difference was observed within the time until noticed between the AOIs of the homepage  $F(5, 180) = 16.205, p = .000$ . The post hoc analysis demonstrated that the meantime, until notice of AOI1, was only significantly lower than that of AOI3 (mean difference = -4.18), AOI4 (mean difference = -7.64), AOI5 (mean difference = -8.04), and AOI6 (mean difference = -7.69) of the home page.

The mean time until noticed of AOI2 was only significantly lower than that of AOI3 (mean difference = -5.44), AOI4 (mean difference = -8.90), AOI5 (mean difference = -9.30), and AOI6 (mean difference = -8.95) of the home page. The mean time until notice of AOI3 was only significantly lower than that of AOI4 (mean difference = -3.46) and AOI5 (mean difference = -3.86).

The analysis result also showed a significant difference within the time viewed between the AOIs of the homepage  $F(5, 180) = 7.244, p = .000$ . The post hoc analysis demonstrated that the mean time viewed of AOI1 was only significantly higher than that of AOI3 (mean difference = 1.38), AOI4 (mean difference = 2.01), and AOI5 (mean difference = 1.95) of the home page. The mean time viewed by AOI2 was only significantly higher than that of AOI3 (mean difference = 1.29), AOI4 (mean difference = 1.92), and AOI5 (mean

difference = 1.86) of the home page. Finally, the mean time viewed by AOI6 was not significantly higher or lower than any other AOIs on the homepage.

AOI 1 and AOI 2 have the highest number of gaze points (AOI 1 = 1301; AOI 2 = 1229), the shortest time to first fixation (AOI 1 = 1.66; AOI 2 = 0.40), and the longest dwell time (AOI 1 = 2.95; AOI 2 = 2.76), suggesting that they attract the most visual attention and all of the important information for visitors should be placed there.

A significant difference was observed within the number of visits between the AOIs of the homepage  $F(5, 180) = 9.082, p = .000$ . The post hoc analysis showed that the total number of visits of AOI1 was only significantly higher than that of AOI 4 (mean difference = 3.13), AOI 5 (mean difference = 2.80), and AOI 6 (mean difference = 3.77) of the home page. The mean number of visits of AOI2 was only significantly higher than that of AOI 3 (mean difference = 3.23), AOI 4 (mean difference = 4.43), AOI 5 (mean difference = 4.10), and AOI6 (mean difference = 5.07) of the home page.

No significant difference was observed in the fixating percentage between the AOIs of the homepage ( $F(5, 180) = 0.896, p = 0.519$ ). While all of the AOIs had high fixating percentages (AOI 1 = 96.67%; AOI 2 = 100%; AOI 3 = 96.67%; AOI 4 = 83.33%; AOI 5 = 90%, AOI 6 = 60%), only AOI 1 (188) and AOI 2 (227) had the highest number of visits suggesting again that these AOIs attracted participants' visual attention the most.

2) *Facial coding*: Tracking of facial expressions elicited during the exposure to the website's homepage was used to pinpoint the emotional activity of the website.

The results suggest that participants mostly elicited sadness (29.55%), neutrality (33.19%), and puzzlement (13.60%) while browsing the homepage, regardless of the AOI (Fig. 3). They also elicited slight disgust (4.33%), fear (3.51%), joy (5.21%), and surprise (29.55%). When combined with survey responses, such emotions can be attributed to the colour scheme used, difficulties in reading the material due to inadequate text colour contrast, and a mismatch between presented information on the homepage and participants' motivation for visiting the page.

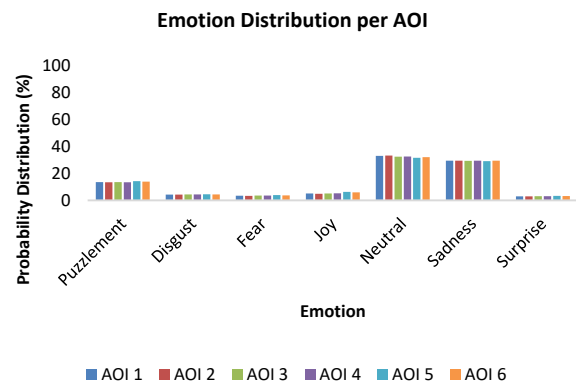


Fig. 3. Probability distribution of elicited emotions per AOI on the OBC's homepage.

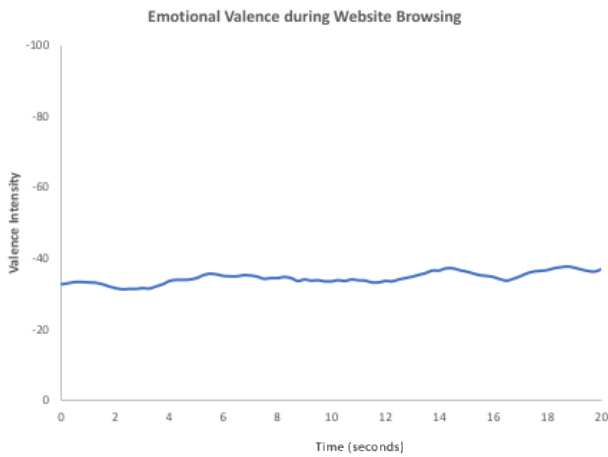


Fig. 4. Emotional valence intensity during OBC’s homepage browsing.

During the 20-second interval of scrolling the homepage, the chart suggests that participants seemed to subconsciously show aversiveness towards the material presented on the website (Fig. 4). A slight decrease in aversiveness was observed between the second and fourth minute as, in that period, participants were looking at the images with the muted colour scheme. With the slight increase of aversiveness as time passed, it is possible that participants did not find the information they wanted and that the colour scheme used created a sensory overload.

Throughout homepage browsing, participants had relatively higher intensities of negative emotions than positive ones (Fig. 5). The most significant difference was observed towards the end of browsing. Scattered gaze points across the homepage in that period could be attributed to the fact that: 1) after the 12th second, nothing was interesting enough to capture their attention; 2) the page layout did not meet their expectations, and they did not find the information they were looking for; 3) the colour scheme used was overly stimulating to hold their attention for a more extended period.

A slight increase in the positive mood intensity was observed when participants looked at the areas with only images with muted colour schemes or used brighter colours to highlight crucial textual information such as upcoming events and student blogs (AOI 4 & AOI 5).

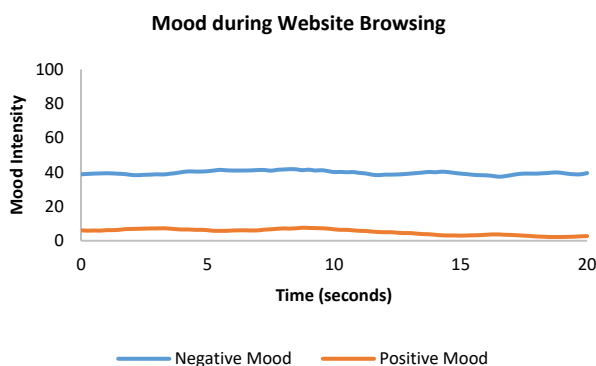


Fig. 5. Intensity of positive and negative emotions while browsing OBC’s homepage.

3) *Survey*: Participants had different reasons for visiting the website: to learn more about different courses (25.5%), to get updates on events (21.8%), to read interesting news (16.4%), and to learn more about research opportunities (18.2%). Several participants (18.2%) also had other (unknown) reasons. While 34.1% of the participants said it was neither hard nor easy to navigate the OBC website, 25% found it somewhat easy, and 20.5% very easy. However, 15.9 % found it somewhat hard, and 4.5% found it hard (Table III).

Regarding the preferences on the website, 45.7% would not change anything, while 42.9% would change the design, colours used, and text. A few participants would change everything (2.9%), while 8.6 % needed to know what they would like to change (Table III). Some of the comments regarding changes were that the text colour on the slider makes the text unreadable, change text colour or background, the website does not look like a college website because of too many colours, and add courses on the homepage.

Most participants did not know the answer (81.6%). A few (15.8%) gave the wrong answer, while only 2.6% answered correctly. Taken together with the heat map, this result suggests that this vital information was not encoded despite being seen.

**B. CARE Page**

The heat map for the CARE page reveals that the top of the CARE page was a point of attraction for participants (Fig. 6). The seen order indicates that participants first notice the page’s heading (AOI 1); however, the most significant point of attraction was the upper part of AOI 2, which is located in the middle of their screen and their eye level height. In contrast, the area dedicated to research resources was not seen at all (AOI 3).

TABLE III. WEBSITE VISITS, NAVIGATION EASINESS, AND PREFERENCES

Participants’ reasons for visiting the website	
Learn more about different courses	25.5%
Event updates	21.8%
News	16.4%
Research opportunities	18.2%
Other	18.2%
OBC website navigation level of easiness	
Hard	4.5%
Somewhat hard	15.9%
Neither hard nor easy	34.1%
Somewhat easy	4.5%
Very easy	20.5%
Preferences	
Change nothing	45.7%
Change (design, colour, text)	42.9%
Change everything	2.9%
Unsure	8.6%



Fig. 6. Heat map and seen order for the OBC's CARE page.

Results also reveal that in this text-based webpage, participants used the 'F' pattern scanning of text where the first few words on the left of each line at the top of the page receive more fixations than subsequent words on the same line, while text towards the end is not seen.

The CARE page was divided into three AOIs to obtain quantitative eye-tracking metrics (Fig. 7). A one-way within-groups analysis of variance (ANOVA) was used to assess the impact of AOIs of the OBC's CARE page on the attention metrics (i.e., Number of fixations, time to the first fixation, time viewed, Number of visits, Fixating Percentage) (Table IV, Table V).

Null hypothesis: There are no significant differences between the number of fixations, Number of visits, time to the first fixation, time viewed, and the fixating percentage between AOIs of the CARE page.

Alternate hypothesis: There are significant differences within the Number of fixations, Number of visits, time to the first fixation, time viewed, and the fixating percentage between AOIs of the CARE page.

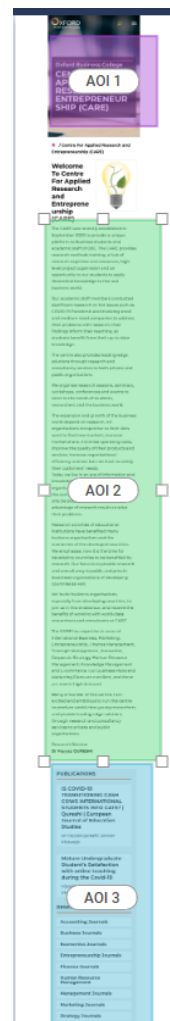


Fig. 7. AOIs for CARE page.

TABLE IV. SUMMARISED STATISTICS FOR CARE PAGE

Number of Fixations		F (2, 180) = 21.347, p = .000	
AOIs		Mean Difference	p-value
AOI 1	AOI 2	-128.43	.000*
	AOI 3	-9.33	.869
AOI 2	AOI 3	119.10	.000*
Number of Visits		F (2, 180) = 29.372, p = .000	
AOI 1	AOI 2	-10.80	.000*
	AOI 3	.13	.995
AOI 2	AOI 3	10.93	.000*
Time to First Fixation		F (2, 180) = 33.249, p = .000	
AOI 1	AOI 2	-2.48	.243
	AOI 3	-11.72	.000*
AOI 2	AOI 3	-9.24	.000*
Time Viewed		F (2, 180) = 22.024, p = .000	
AOI 1	AOI 2	-8.60	.000*
	AOI 3	-.63	.860
AOI 2	AOI 3	7.96	.000*
Fixating Percentage		F (2, 180) = 0.171, p = .848	

TABLE V. AVERAGE VALUES OF EYE-TRACKING METRICS FOR THE CARE PAGE

Time to First Fixation (s)	
AOI	M
AOI 1	0.16
AOI 2	2.63
AOI 3	15.49
Number of Fixations	
AOI 1	1715
AOI 2	5568
AOI 3	1995
Number of Visits	
AOI 1	148
AOI 2	472
AOI 3	144
Time Viewed (s)	
AOI 1	3.77
AOI 2	12.37
AOI 3	5.75
Fixating percentage (%)	
AOI 1	100
AOI 2	100
AOI 3	76.67

A significant difference was observed within the number of fixations between the AOIs of the CARE page (F (2, 180) = 21.347, p = .000), where the total number of fixations of AOI 2

was significantly higher than that of AOI 1 (mean difference = 128.43) and AOI 3 (mean difference = 119.10)). The total number of fixations of AOI 3 was significantly lower than that of AOI 2 (mean difference = -119.10).

A significant difference was observed within the time to first fixation between the AOIs of the CARE page (F (2, 180) = 33.249, p = .000), where the average time to first fixation of AOI 1 was only significantly lower than that of AOI 3 (mean difference = -11.71) and the average time to first fixation of AOI 2 was significantly lower than that of AOI 3 (mean difference = -9.24).

A significant difference was also observed for the time spent looking at a particular AOI (F (2, 180) = 22.024, p = .000), where the average time viewed of AOI 2 was significantly higher than AOI 1 (mean difference = 8.59) and AOI 3 (mean difference = 7.96). No significant difference in the average time viewed was observed between AOI 1 and AOI 3.

Taken together, AOI 2 has the highest number of gaze points (5568) and the longest dwell time (12.37 s) when compared to AOI 1 (1715; 3.77 s) and AOI 3 (1995; 5.75 s), suggesting that it attracts the most visual attention. Since AOI 1 has the shortest time to first fixation (0.16 s), followed by AOI 2 (2.63 s) and AOI 3 (15.49 s), the information about research resources from AOI 3, to be seen, should be placed between AOI 1 and AOI 2, as it is the vital point of the CARE.

A significant difference was observed within the total number of visits between the AOIs of the CARE page (F (2, 180) = 29.372, p = .000). The post hoc analysis demonstrated that the total number of visits of AOI 2 was significantly higher than that of AOI 1 (mean difference = 10.80) and AOI 3 (mean difference = 10.93). In contrast, no difference was observed in the total number of visits between AOI 1 and AOI 3.

No significant difference in the fixating percentage was observed (F (2, 180) = 0.171, p = .848).

#### IV. DISCUSSION

Eye-tracking behaviour neurometrics from OBC's homepage suggest that the top parts of the webpage are noticeable, especially those in eye-level height. Since gaze points were concentrated on the webpage's upper left and center positions, it is suggested to position all the vital information regarding offered courses, events, and research opportunities (Fig. 4) there. In that way, not only will visitors' attention be grabbed, but most importantly, they will also be given the requested information and will even be motivated to scroll down the page afterward.

A comparison between the homepage and the CARE page suggests that the time to the first fixation, the number of fixations, the number of visits, and the time viewed highly depend on the visited page. In contrast, the fixation percentage is the only observed metric varying in the same range regardless of the visited page (Table VI). Moreover, all the metrics were higher for the CARE page, indicating that the latter requires higher visual attention than the homepage.

TABLE VI. COMPARATIVE TABLE FOR OBC'S HOMEPAGE AND CARE PAGE

M (OBC's homepage)	M (CARE page)
<i>Time to First Fixation (s)</i>	
0.40-9.70	15.49-0.16
<i>Number of Fixations</i>	
1301-746	5568-1715
<i>Number of Visits</i>	
277-75	472-144
<i>Time Viewed (s)</i>	
2.95-1.00	12.37-3.77
<i>Fixating percentage (%)</i>	
100-60	100-76.67

Observed sadness, puzzlement, and neutrality regardless of the AOI within the homepage, together with survey answers, could be attributed to the colour scheme used (overly stimulating), difficulties reading the material due to inadequate colour-text contrast, and mismatch between the order of presented information on the homepage and participants reasons for visiting the page. This was also supported by an observed increase in the intensity of positive mood and a decrease in the aversiveness in the page areas where images with the muted colour scheme were used. In the future, it is suggested to use neutral colours for backdrops and one or two stimulating, bright colours to highlight materials of interest [19].

Based on the eye-tracking behavioural neurometrics, emotion analysis, and survey answers, a re-design is recommended as information of interest for visitors needs to be optimally positioned on the website, and most of it goes unnoticed. Even if noticed, it does not get encoded and recalled due to inappropriate colour-text contrast(as seen in the survey on slide 13) [20].

## V. CONCLUSION

The findings of this research showcased the importance of college websites and their crucial role in attracting regional and international students. Colleges must build better, more user-friendly websites as current and prospective students increasingly rely on technology. The gaze activity was assessed over time to determine their attention distribution. The heat maps were essential in observing the movement pattern around the website. Emotion analyses using facial coding to gain insight into their behaviour and reactions to the website played an imperative role during this website testing. The results showed that participants' negative feelings were more intense than their good emotions when the content required page scrolling, while a difference was noticed near the end of the browsing session. However, looking at regions with just photos in a subdued colour scheme or where brighter colours were used to emphasize essential textual information such as forthcoming events and student blogs, the participants' pleasant mood intensity increased somewhat. Lastly, participants employed the 'F' pattern monitoring of text on the text-based homepage, where the first few words on the left of

each line at the top of the page received more fixations than other words on the same line, while text near the end was not viewed. According to these lines, this research demonstrates how with the help of neuroscientific research tools, an insight into students' behaviour and their subconscious preferences. Despite the use of online study with the webcam-based eye tracking solution (15Hz), the obtained results are an example of how very accurate and valuable marketing insights of the consumer's subconscious on the website's perceived process and preferences can be obtained. The limitation of the study was the big screenout (92 subjects) and 124 participants who partially were involved in the study (participants who started the experiment but closed the browser or had timed out before reaching the end), where we presume this is due to the reason what participants did not receive any incentives for attending to this research. In addition, the software used is susceptible and discredits the participant at a slight shift of the head, which gives us merit results at the very end. Since the research was limited to the mobile device based on the website stats of visitors per device, to achieve the best results possible, the total sample size needed to be increased on other devices, too, which will result in a better percentage of quality recordings.

### A. Recommendations

In this age, where everything relies on technology, providing students with appropriate websites they can depend on to find the needed information is necessary, as they often need an alternative. Both homepage and CARE required a certain amount of attention with varying magnitude depending on the website's content.

### B. Homepage

For this case, enhancing the user's experience by:

- keeping the text clear and concise and avoiding complex language.
- using bright colours.
- using text alternatives such as images, videos could be used.

### C. CARE

The higher metrics (Number of fixations, time to the first fixation, time viewed, number of visits) compared to the homepage reveal that further action should be taken to reduce the page complexity and the cognitive load and to make it more accessible to students. Such action includes changing the page's design, colour, and text since over 42% opted for such changes.

Similar suggestions to the homepage can enhance the number of fixations, and time to the first fixation, time viewed. Moreover, the higher number of visits can be improved by providing ways to help the visitor get around the website, identify the main content, and where the navigation is.

## ACKNOWLEDGMENTS

This study was done in collaboration with the Institute for Neuromarketing, Zagreb, Croatia, which provided support for neuromarketing equipment and data analysis. The research was conducted exclusively for development and scientific advancement. The authors would like to acknowledge the

support of Oxford Business College in paying for the Article Processing Charges (APC) for this publication and the Institute for Neuromarketing for covering all neuromarketing research costs.

### REFERENCES

[1] Poock, M. C., & LeFond, D. (2001). How college-bound prospects perceive university websites: Findings, implications, and turning browsers into applicants: college and University, 77(1), 15.

[2] Schimmel, K., Motley, D., Racic, S., Marco, G. and Eschenfelder, M. (2010). 'The importance of university web pages in selecting a higher education institution', Research in Higher Education Journal, Vol.9, 1.

[3] Alstete, J.W. and Beutell, N.J. (2004) 'Performance indicators in online distance learning courses: a study of management education, Quality Assurance in Education, Vol.12, pp.6-14 doi:10.1108/09684880410517397.

[4] F. Espigares-Juradoa, F. Muñoz-Leiva, M. B.Correia, C. M. R. Sousad, C. M. Q.Ramose, L. Faiscaf, 'Visual attention to the main image of a hotel website based on its position, type of navigation and belonging to Millennial generation: An eye tracking study', Journal of Retailing and Consumer Services, vol 52, January 2020.

[5] I. Michael, T. Ramsay, Me. Stephens, F. Kotsi, 'A study of unconscious emotional and cognitive responses to tourism images using a neuroscience method', Journal of Islamic Marketing, January 2019.

[6] Y. M. Hwang, K. C. Lee, 'Using an Eye-Tracking Approach to Explore Gender Differences in Visual Attention and Shopping Attitudes in an Online Shopping Environment', International Journal of Human-Computer Interaction vol 34, Issue 1, April 2017, pp.15–24.

[7] E. A. V. Reijmersdal, E. Rozendaal, L. Hudders, I. Vanwesenbeeck, V.Caubergh, Z. M.C. V. Berlo, 'Effects of Disclosing Influencer Marketing in Videos: An Eye Tracking Study among Children in Early Adolescence' Journal of Interactive Marketing, vol 49, Issue 1, February 2022.

[8] K. Keib, C. Espina, Y.I Lee, B. W. Wojdyski, D. Choi, H. Bang, 'Picture This: The Influence of Emotionally Valenced Images, On Attention, Selection, and Sharing of Social Media News' Media Psychology, vol 21, October 2017, pp 202–221.

[9] A. Duque, C. Vázquez, 'Double attention bias for positive and negative emotional faces in clinical depression: Evidence from an eye-tracking

study' Journal of Behavior Therapy and Experimental Psychiatry, vol 46, March 2015, pp 107–114.

[10] J. Zhang, A. B. Chan, E. Y. Y. Lau, J. H. Hsiao, 'Individuals with insomnia misrecognize angry faces as fearful faces while missing the eyes: an eye-tracking study', Sleep, vol 42, Issue 2, February 2019.

[11] U. Bhandaria, K. Changa, T. Nebenb, 'Understanding the impact of perceived visual aesthetics on user evaluations: An emotional perspective', Information & Management, vol 56, Issue 1, January 2019, pp 85–793.

[12] A. T. Stulla, L. Fiorellab, R. E.Mayera, 'An eye-tracking analysis of instructor presence in video lectures', Computers in Human Behavior, vol 88, November 2018, pp 263–272.

[13] P. Klein, J. Viiri, S. Mozaffari, A. Dengel, J. Kuhn, 'Instruction-based clinical eye-tracking study on the visual interpretation of divergence: How do students look at vector field plots?' Phys. Rev. Phys. Educ. Res. vol 14, March 2018.

[14] M. Azimzade, F. Nobakht. Z. Aminiroshan, M. S. Sotoode, 'Employing Eye Tracking in Quantifying and Qualifying Visual Attention of Web Site Viewers (Physical Education Faculties)', Journal of Advanced Sports Technology, vol 4, Issue 1, June 2020, pp 9–19resources/publications/ethicalguidelines-for-educational-research-2018/.

[15] Mestre, L.S. (2012). ' Assessment of learning objects. Designing Effective Library Tutorials, pp.205–201, <https://doi.org/10.1016/B978-1-84334-688-3.50009-3>.

[16] Sungkur, R. K., Antoaroo, M. A. and Beeharry, A. (2016). 'Eye tracking system for enhanced learning experiences, Education and Information Technologies, Vol.21, pp.1785-1806 doi:10.1007/s10639-015-9418-0.

[17] Faul, F., Erdfelder, E., Lang, A.-G. and Buchner, A. (2007). 'G\* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences, Behavior research methods, Vol.39, pp.175-191 doi:10.3758/BF03193146.

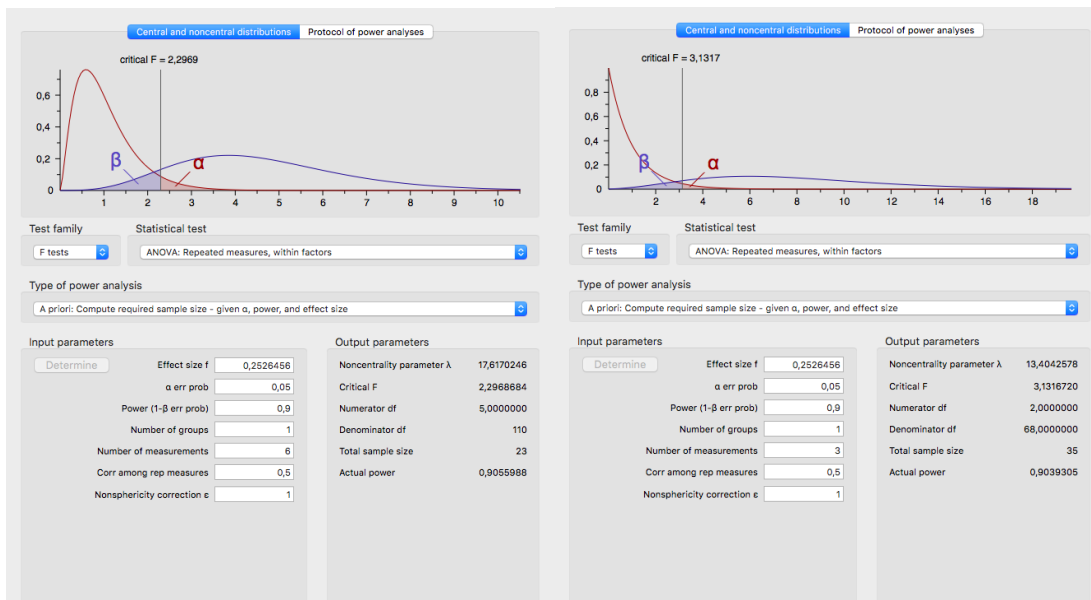
[18] British Educational Research Association [BERA] (2018), Ethical Guidelines for Educational Research, fourth edition, London. <https://www.bera.ac.uk/researchers>.

[19] Wenham, D. and Zaphiris, P. (2003). 'User interface evaluation methods for internet banking websites: a review, evaluation and case study, Human-Computer Interaction, Theory and Practice, pp.721-725.

[20] Heimlich, J.E. and Wang, K. (1999) Evaluating the structure of web sites.

### APPENDIX

G\*Power output for OBC homepage and CARE page.



Based on the G\*power output [16], to detect the effect with 90% power and a two-sided significance level of 5%, a sample size of n=23 for the OBC homepage and n=35 for the CARE page is required.

Survey questions used in the study:

- 1) How often do you visit this website?
- 2) Never / Very Rarely (once per month) / Rarely (2-3 times per month) / Occasionally (2-3 times per week) / Frequently (1-2 times per day) / Very Frequently (more than three times per day)
- 3) Why are you visiting this website?  
To learn more about the offered courses / To read interesting news / To get updates on events / To learn about research opportunities and resources / Other (please specify)
- 4) How easy was it to navigate this website?  
Very Hard / Somewhat Hard / Neither Hard nor Easy / Somewhat Easy / Very Easy
- 5) What would you like to change on this website?
- 6) What is the name of the recent award that OBC got?



# Improving MapReduce Speculative Executions with Global Snapshots

Ebenezer Komla Gavua<sup>1</sup>, Gabor Kecskemeti<sup>2</sup>

Institute of Information Technology, Miskolc-Egyetemvaros 3515, Miskolc, Hungary  
Computer Science Department, Koforidua Technical University, Koforidua, Ghana<sup>1,2</sup>

Department of Computer Science  
Liverpool John Moores University, Liverpool, UK<sup>2</sup>

**Abstract**—Hadoop’s MapReduce implementation has been employed for distributed storage and computation. Although efficient for parallelizing large-scale data processing, the challenge of handling poor-performing jobs persists. Hadoop does not fix straggler tasks but instead launches equivalent tasks (also called a *backup task*). This process is called Speculative Execution in Hadoop. Current speculative execution approaches face challenges like incorrect estimation of tasks run times, high consumption of system resources and inappropriate selection of backup tasks. In this paper, we propose a new speculative execution approach, which determines task run times with consistent global snapshots and K-Means clustering. Task run times are captured during data processing. Two categories of tasks (i.e. fast and stragglers) are detected with K-Means clustering. A silhouette score is applied as decision tool to determine when to process backup tasks, and to prevent extra iterations of K-Means. This helped to reduce the overhead incurred in applying our approach. We evaluated our approach on different data centre configurations with two objectives: *i*) the overheads caused by implementing our approach and *ii*) job performance improvements. Our results showed that *i*) the overheads caused by applying our approach is becoming more negligible as data centre sizes increase. The overheads reduced by 1.9%, 1.5% and 1.3% (comparatively) as the size of the data centre and the task run times increased, *ii*) longer mapper tasks runs have better chances for improvements, regardless of the amount of straggler tasks. The graphs of the longer mappers were below 10% relative to the disruptions introduced. This showed that the effects of the disruptions were reduced and became more negligible, while there was more improvement in job performance.

**Keywords**—MapReduce; Hadoop; speculative executions; stragglers; consistent global snapshots; K-means algorithm

## I. INTRODUCTION

The Hadoop software environment provides a widespread implementation for distributed data storage and MapReduce computing [1], [2], [3]. However, the challenge of handling poor-performing jobs persists. Hadoop launches equivalent tasks (also called a *backup task*) in place of straggler tasks to finish the computation faster. This process is Hadoop’s speculative execution [4].

Previous research into speculative execution has shown efforts to improve job performance in MapReduce. Past strategies such as LATE [5] and MCP [6] recognize straggler tasks based on self-estimation of the tasks’ remaining time. SAMR [7] and ESAMR [8] use historical information to classify nodes into slow map and reduce nodes. SECDT [9] predicts the remaining time of running tasks based on real-time information on tasks. However, these previous approaches

have various problems. Most have challenges with the accurate estimation of the remaining time of slow tasks. Some have significant overheads during the estimation of straggler tasks’ remaining times.

Our work proposes a new speculative execution approach, which estimates task runtimes with consistent global snapshots and K-Means clustering. Task progress is captured consistently during data processing. Two categories of tasks (fast and straggler) are identified with K-Means. A silhouette score is applied as decision tool to determine when to process backup tasks. This helped to reduce the overhead incurred in applying our approach since, the stragglers were quickly detected and rescheduled.

We evaluated our approach on different data centre configurations. The data centre configurations were selected after considering Hadoop cluster requirements from the industry. We focused our experiments on two objectives: *i*) the overheads caused by implementing our approach and *ii*) job performance improvements. Two categories of backup tasks were considered in our experiments. *i*) backup tasks that can transfer their states when shifting from one node to another and *ii*) backup tasks that need to restart after their transfer (i.e. tasks supported by Hadoop’s original speculative execution).

We experimented with mappers as there are typically more of them in a Hadoop application than reducers. Since reducer handling is done the same way as mapper handling in Hadoop, our results are applicable to both. We also focused on different durations of mappers, which provided the details of how long the mappers took to process data. From these, we concluded on the following *i*) the overheads caused by applying our approach is becoming more negligible as data centre sizes increase. The overheads reduced by 1.9%, 1.5% and 1.3% (comparatively) as the size of the data centre and the task run times increased. *ii*) longer mapper tasks runs have better chances for improvements, regardless of the amount of straggler tasks. The graphs of the longer mappers were below 10% relative to the disruptions introduced. This showed that the effects of the disruptions were reduced and became more negligible, while there was more improvement in job performance.

The remainder of this paper is structured as follows. In Section II, we reviewed concepts and related works about MapReduce and Speculative Executions. In Section III, we present our methodology for detecting straggler tasks and proposed improvements. In Section IV, we discussed the

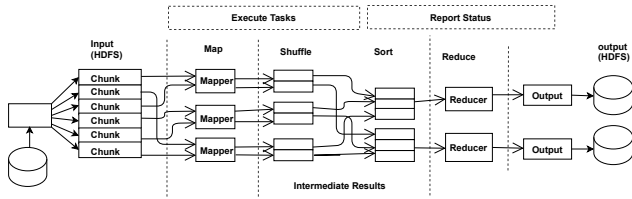


Fig. 1. MapReduce structure.

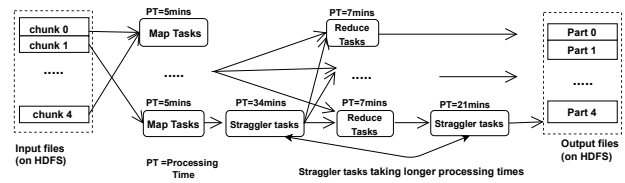


Fig. 2. Speculative execution.

procedure for evaluating our approach via experiments and we analysed the results. Section V concludes the paper with recommendations for future work.

## II. CONCEPTS AND RELATED WORKS

### A. Concepts

In this subsection, we briefly discuss few concepts related to the MapReduce programming model.

1) *MapReduce*: MapReduce (MR) is a programming model for massive data computing used in Apache Hadoop. MapReduce is used for writing applications that process and analyse large data sets. These applications ran in a parallel fashion on large clusters in a scalable and fault-tolerant manner. A MapReduce job breaks and divides the input data into chunks which are first processed by the “Map phase” in parallel and then by the “Reduce phase” [10], [11] as seen in Fig. 1.

2) *Hadoop*: Apache Hadoop is an open-source software implementation of MapReduce. The core of Hadoop includes a distributed file system, and a MapReduce processor [1]. The Hadoop distributed file system (HDFS) works closely with MapReduce by distributing storage and computation across large clusters [12], [13]. During job processing on Hadoop, if a task of a job requires an abnormally long execution time, the total completion time of the job is affected. Such a task is called a straggler task. MR reruns straggler tasks on a different machine to finish the computation faster. The process of diagnosing straggler tasks and assigning them to other nodes is called *speculative execution* [14]. These faults are mainly due to IO contentions, background services, hardware behaviours, unbalanced load or uneven distribution of resources and other reasons [15]. Some straggler tasks run significantly slower than other tasks as shown in Fig. 2, where the straggler tasks take more processing times than the normal MR tasks [4].

3) *Prior speculative execution strategies*: A couple of research activities have been conducted to solve poor-performing tasks.

The Hadoop Naïve Method was implemented with the Hadoop architecture. However, most of the tasks processed during runtime were detected as slow tasks and processed as backup tasks. This affected job completion because there was no improvement in job completion time after processing the backup tasks. Also, this strategy is not suitable in heterogeneous environments. Therefore, an approach that distinguishes straggler tasks from the normal tasks during job processing will ensure job performance improvements.

Zaharia et al. [5] developed the Longest Approximate Time To End (LATE) algorithm. LATE is a simple, robust scheduling

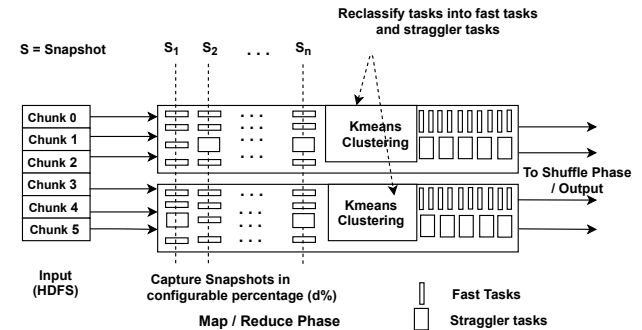


Fig. 3. Structure of task performance monitoring algorithm.

algorithm that uses estimated finish times to detect straggler tasks. LATE is not suitable in heterogeneous environments. Therefore, a dynamic approach that works in all types of environments will help estimate the task runtime to ensure the improvement of job performance.

Chen et al. [7] proposed a Self-Adaptive MR Scheduling Algorithm (SAMR). SAMR uses historical information to classify nodes into the slow map- and reduce-nodes. This makes SAMR dependent on previous tasks information. Therefore, an approach that applies the information of current tasks without depending on previous nodes will be welcomed in the research community.

Sun et al. [8] designed an Enhanced Self-Adaptive MR Scheduling Algorithm (ESAMR) as an improvement on SAMR by utilising the K-means clustering algorithm to classify historical information. Therefore, the reliance of ESAMR on previous task information makes it only applicable when there is historical information. Moreover, the K-means clustering algorithm utilised was not validated to determine the straggler tasks. Therefore, an approach that is not affected by changes in dataset and validates the kmeans clustering will allow users to better assess tasks behaviours.

Chen et al. [16] proposed the Maximum Cost Performance approach, which considers the cost performance of cluster computing resources to estimate the slow tasks. However, in the map phase, task satisfying data localisation executes faster than those not satisfying data localisation. This provides an unfair comparison between the tasks at the same level. Therefore, an approach that considers all tasks at the same level will ensure an appropriate estimation of task run times.

Huang et al. [9] proposed a new Speculative Execution Algorithm based on C4.5 Decision Tree (SECDT) to improve predicted execution times among previous research resulting in poor job performance. However, navigating the decision tree implemented by this strategy is prone to significant

overheads. Therefore, an approach that determines task run times via snapshot captures will enable the improvement of job performance.

In summary, the existing speculative execution strategies still encounter challenges in managing straggler tasks in Hadoop. We now discuss the design of our proposed approach.

### III. METHODOLOGY

This section focuses on the design of our approach which is designed to improve job performance on MapReduce Hadoop (MRH). The approach consists of two algorithms that are interconnected to ensure correct determination of task run times, appropriate selection of backup tasks and reduction in the consumption of system resources. The goals of this section are:

- To design an algorithm that captures task run times during data processing on mappers and reducers. This is achieved by repetitive capturing of the task run times at specific intervals.
- To design an algorithm that monitors task performance on their nodes to foster the rescheduling of straggler tasks to available nodes for reprocessing.
- To implement K-means clustering algorithm to determine straggler tasks. The K-means clustering algorithm is applied with the Silhouette Coefficient to validate the outputs of the clustered data sets.
- To assess the algorithms on scalable configurations of MRH to prove their applicability. A survey of industry and real-life MRH configurations is conducted to ensure that the solution is applicable in industry.

#### A. Consistent Global Snapshots on MapReduce

This approach comprises of snapshot capturing and task performance monitoring algorithms as seen in Algorithms 1 to 2, and Fig. 3 to 5. Fig. 4 shows state transitions during the capturing of task run times. The two algorithms work together to ensure that straggler tasks are detected correctly and processed as backup tasks.

This approach is designed to dynamically collect real-time data from all types of environments. The collected real-time data fosters the early detection of straggler tasks to reduce high consumption of system resources. Moreover, this approach is applicable in most environments compared to a few existing approaches which struggle in heterogeneous environments. Additionally, some of the existing approaches have limitations with accurate estimation of remaining time of straggler tasks. Also, some have significant overheads during the estimation of straggler tasks' remaining times. The details of the algorithms are discussed below.

1) *Snapshots capturing algorithm*: Algorithm 1 is applied to explain the snapshots capturing process. The algorithm is designed with specific parameters to foster its comprehension.  $T_s$  and  $C_s$  are utilised to model tasks state transitions (*Task\_state*) and the snapshot capturing state transitions (*Snap\_state*) during job processing as seen in Fig. 5 and 4. These two parameters

---

#### Algorithm 1 Snapshot Capturing Algorithm

---

**Require: Variables:**  $T_s = Task\_state, C_s = Snap\_state,$   
 $N = Node$   
**Require: Variables:**  $Q_I = Tasks\_Instances,$   
 $i = counter\ for\ Q_I$   
**Require: Variables:**  $G_s = captured\_snapshots$   
**Require: SnapStateFunction:**  
 $SnapState : C_s \rightarrow \{snap\_ready,$   
 $snapping, snap\_paused, snap\_completed\}$   
1: **for**  $i < Q_I$  **do**  
2:   **if**  $T_s = ready$  **then**  
3:      $C_s := snap\_ready$  // Snap\_state updates to ready  
      when tasks processing begins.  
4:   **else**  
5:      $C_s := pause\_snapping$   
6:     check the system and restart the task  
7:   **end if**  
8:   **while**  $T_s = running$  **do**  
9:      $C_s := snapping$  // Snap\_state updates to snapping  
      during tasks processing.  
10:     Save the captured snapshots  
11:      $G_s = G_s + 1snapping$   
12:   **end while**  
13:    $i = i + 1$  // Tasks are monitored until they are  
      processed.  
14: **end for**  
15: **if**  $T_s = terminated$  **then**  
16:    $C_s := pause\_snapping$  // Snap\_state pauses when a  
      task is terminated with task\_state updated to terminated.  
17: **end if**  
18: **while**  $i \leq Q_I$  **do**  
19:    $C_s := snapping$  // Task run times capturing continues  
      until all tasks are processed.  
20:   Save the captured snapshots  
21:   **if**  $T_s = completed$  **then**  
22:      $C_s := snapping\_completed$  // Snap\_state updates  
      to snap\_completed when tasks processing ends.  
23:     Save the captured snapshots  
24:      $G_s = G_s + 1snapping$   
25:   **end if**  
26:    $i = i + 1$   
27: **end while**

---

together with the *SnapStateFunction* and the *TaskStateFunction* help to describe the status of task processing and snapshot capturing at any specific period.

Algorithm 1 initialises with task processing to foster the capturing of task run times as seen in Fig. 3. Nodes ( $N$ ) are monitored before task processing begins. This is done to capture the commencement of task processing (*start times*), as seen in the snapshot capturing state diagram in Fig. 4. When job processing commences, data is uploaded into the system for task processing to commence.

*SnapStateFunction* is activated, which causes *Snap\_state* ( $C_s$ ) to be updated to *ready\_snap* as seen in lines 1 to 3 of Algorithm 1. However, if a task is not ready (due to a fault),  $C_s$  is updated to *pause\_snapping* and the system is checked (for the task to be restarted) as seen in lines 4 to 7.

While the updated data is being processed,  $C_s$  is updated

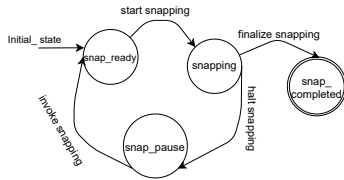


Fig. 4. Snapshot capturing state transition diagram.

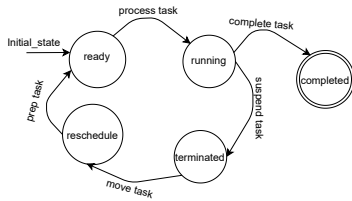


Fig. 5. Tasks state transition diagram.

from *snap\_ready* to *start\_snapping* as as seen in lines 8 to 13. Task run times are captured and saved as seen lines 9 to 10. Task run times are captured on all nodes and saved on snapshots text files. Snapshots are captured repeatedly for all the tasks running on nodes. The accumulated local snapshots captured constitute the global snapshots ( $G_s$ ). The details captured include (i) task start-time, (ii) task completion-time, (iii) the node ( $N$ ) on which the task is running, (iv) task identification, and (v) task status. The task start-time is the specific time the processing of a particular task  $t$  commences. While the completion-time is the specific time the processing of  $t$  ends. The task identification is the unique key given to every task  $t$  when their processing commences. The task status is reflective of the current state of  $t$  as seen in the state transition Fig. 5.

Additionally, the *TaskStateFunction* is activated, which causes *task\_state* ( $T_s$ ) to be updated from *ready* to *running* as seen in the task state diagram in Fig. 5.  $T_s$  remains unchanged until all the tasks are completely processed. Then, it transitions from *running* to *completed*. However, when a task's run time is unnecessarily longer that expected, the task is suspended, which causes  $T_s$  to be updated to *terminated* and  $C_s$  to *pause\_snapping* as seen in lines 15 to 17.

When a configurable percentage of the tasks have been processed with captured run times as seen in Fig. 3; K-means clustering algorithm is employed to classify the captured data on the snapshots text files, to determine the straggler tasks. The straggler tasks identified are then processed as backup task on available nodes. This causes  $T_s$  to be updated from *terminated* to *rescheduled* as seen in Fig. 5.

When the tasks rescheduling is completed, the re-processing of the backup tasks commences. This causes  $T_s$  to transition from *rescheduled* to *ready*. The backup tasks are processed together with the snapshot capturing until all the tasks are completely processed as seen in lines 18 to 27 of Algorithm 1.

2) *Task performance monitoring algorithm*: The task performance monitoring and the snapshot capturing algorithms work concurrently to ensure job performance improvement, as

### Algorithm 2 Task Performance Monitoring Algorithm

**Require: Variables:**  $Q_I = Tasks\_Instances$ ,  
 $T_R = running\ tasks$ ,  $i = counter\ for\ Q_I$   
**Require: Variables:**  $N_i = Node$ ,  $A_n = AvailNodes$ ,  
 $T_C = completed\ tasks$   
**Require: Variables:**  $Q_I = \{t_1, t_2, t_3, \dots, t_n\}$ ,  $T_s = Task\_State$   
**Require: Variables:**  $T_{ET} = task\ execution\ time$ ,  
 $T_{MET} = task\ maximum\ execution\ time$   
**Require: TaskStateFunction:**  
 $TaskState : T_s \rightarrow \{ready, running, terminate, reschedule, completed\}$   
1: *Begin Tasks Processing in the Map or Reduce Phase*  
2: **for**  $i \leq Q_I$  **do**  
3:  $status \leftarrow checkTasksStatus$   
4: **switch** ( $status$ )  
5:  $T_s := ready$  // task is ready for processing.  
6: **case** *still\_Running*:  
7: *monitor the progress of the task*  
8:  $T_s := running$   
9: **if**  $T_{ET} > T_{MET}$  **then**  
10: *terminate the task*  
11:  $T_s := terminated$  // straggler tasks are stopped.  
12: *reschedule straggler tasks on available nodes*  
13:  $T_s := rescheduled$   
14:  $T_R \leftarrow (T_R + t)$  // Running tasks list increased.  
15: **else**  
16: *process all the tasks*  
17: **end if**  
18: **case** *finished\_Running*:  
19: *Tasks completely processed*  
20:  $T_s := completed$   
21: *Output results*  
22:  $A_n \leftarrow (A_n + N_i)$  // Available nodes list increased for backup task.  
23:  $T_C \leftarrow (T_C + t)$  // Monitor the tasks completed.  
24:  $Q_i \leftarrow (Q_i - t)$  // Task instance list is reduced.  
25: **end switch**  
26: **if**  $|Q_i| = 0$  **then**  
27: *Stop tasks monitoring*  
28: **else**  
29: *Continue with tasks monitoring*  
30: **end if**  
31:  $i = i + 1$  // Counter is increased to process all tasks.  
32: **end for**

seen in Fig. 3.

Algorithm 2 is applied during task processing to monitor and evaluate task performance. When data processing begins, all tasks i.e.,  $t = t_1, t_2, t_3, \dots, t_n$  are expected to process data at the same rate. These tasks are allocated processes on compute nodes as seen in lines 1 to 2.

The *TaskStateFunction* is activated which causes  $T_s$  to be updated to *ready* as seen in line 5. Tasks-instances ( $Q_I$ ) are monitored to determine whether they are still running ( $T_R$ ) or are completely processed ( $T_C$ ) as seen in lines 6 and 18.

During task processing, tasks which have relatively longer run times than the maximum execution times ( $T_{MET}$ ) of the

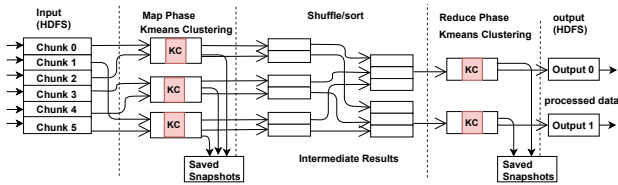


Fig. 6. Structure of algorithm implementation with k-means clustering.

$$J(V) = \sum_{i=1}^{C_d} \sum_{j=1}^{C_i} (|x_i - y_j|)^2 \quad (1)$$

Also, the K-means algorithm is implemented with a validation technique as a decision-making tool in our work. It directs whether to process backup tasks or not. There are cases where the dataset presented for clustering is uniform. However, K-means still tries to cluster it. Thus, clustering results require validation to determine the goodness of fit of the clusters created as seen in Fig. 7.

tasks being processed are terminated as seen in lines 6 to 10. This causes  $T_s$  to be updated to *terminated* as seen in line 11 and Fig. 5. The K-means algorithm is applied to cluster all the captured task run times as seen in Fig. 6 on page 16. The tasks identified as straggler tasks are rescheduled as seen in lines 12 to 14.  $T_s$  transitions from *terminated* through *rescheduled* to *ready* as seen in Fig. 5.

The states of the tasks which do not exhibit relative longer run times, transition from *ready* through *running* to *completed*. This enables their compute nodes to be availed for processing backup tasks, and reduces the number of task instances. These processes are seen in lines 18 to 25.

A vital aspect of this algorithm is the monitoring of task instances ( $Q_I$ ). The number of tasks are monitored throughout their processing stages. When the number of active tasks are exhausted, job processing ends. Otherwise, the task processing continues until the jobs generated are completely processed seen in lines 26 to 32.

### B. Identifying Straggler Tasks with K-Means Clustering Algorithm

The identification of straggler tasks during job processing was a challenge that required addressing in our approach. This was achieved via the adoption of a clustering technique.

Clustering was considered because it is the type of unsupervised machine learning where its goal is to partition sets of objects into groups called *clusters*. These groups can be mutually exclusive or they may overlap, depending on the approach used. It is in contrast to the supervised learning techniques where the goal is to make predictions about output value  $y$  given an input object or instance  $x$  [17]. This made the choice of clustering suitable for our approach since there was no need for training any data set to achieve our groupings.

Additionally, we considered K-means clustering as the clustering technique for our approach because it is a hard clustering algorithm which delivers mutually exclusive groupings. K-means partitions a set of  $n$  objects into  $k$  clusters, so that the resulting intra-cluster similarity is high but the inter-cluster similarity is low [18]. It was the most suitable clustering algorithm for our approach since two distinct groups are required; thus fast tasks and straggler tasks.

Our approach applied the K-means clustering algorithm to categorise task run times (dataset) received from the snapshot capturing algorithm. The dataset saved on snapshots text files during the map or the reduce phases are clustered into fast and straggler tasks as seen in Fig. 6. K-means optimizes the distance between the task run times to their centre points, as seen in eq. (1) [19].

In order to ensure the effective creation of clusters, four clustering validation techniques were considered. These are Dunn [20], Davie-Bouldin, Calinski-Harabasz indices, and the Silhouette score [21]. However, the silhouette score was selected for our approach. The first three were not implemented because of the following: first, although the Calinski-Harabasz index defines how dense and separated a cluster is, the absence of upper- and lower-bounds ranges made it inapplicable. Second, the Davies-Bouldin index utilizes zero (0) as the upper bound; and values closer to zero indicate a better partition. Moreover, the Davies-Bouldin index did not have a lower bound. In the case of the Dunn index, higher indices indicate better clustering. However, the absence of a lower bound makes it inapplicable in our context. Since, without a closed range of clustering validation values, a deterministic algorithm based on them would be unreliable. Also, the presence of the upper-lower bounds fosters faster determination of the goodness of fit of clusters created. Its absence introduces extra overheads into our strategy and makes the choice inappropriate for our research. Nevertheless, the Silhouette score  $S_i$  utilizes an easy-to-evaluate metric to determine the goodness of the clustering. Silhouette score values have a closed range of -1 to 1 [22]. Thus, the silhouette score was chosen for this work.

Algorithm 3 is utilised to identify the suitability of a clustering output for fast and straggler tasks. This algorithm validates the silhouette scores after the clustering exercise. It utilizes the values to decide whether to process backup tasks or not. For instance, Fig. 7 shows two very close data clusters which is difficult to ascertain the fast tasks or poor-performing ones. However, Fig. 8a to 8d on page 17 show well defined data clusters which will require rescheduling of straggler tasks.

#### Algorithm 3 Kmeans Clustering Validation Algorithm

**Require:** Set the  $S_i$  threshold lower – bound as  $Z_x = 0.685$   
**Require:** Set the  $S_i$  threshold upper – bound as  $Z_y = 0.99$

- 1: Initialize the clustering output as an array  $A[k]$
- 2: **for**  $k = 1$  **to**  $A.length$  **do**
- 3:   **if**  $S_i > Z_x$  &  $S_i \leq Z_y$  **then**
- 4:     Reschedule tasks on available nodes
- 5:   **else**
- 6:     Run the tasks on current nodes
- 7:   **end if**
- 8: **end for**

The results from the silhouette score are utilized to determine the goodness of the K-means clustering. If the silhouette

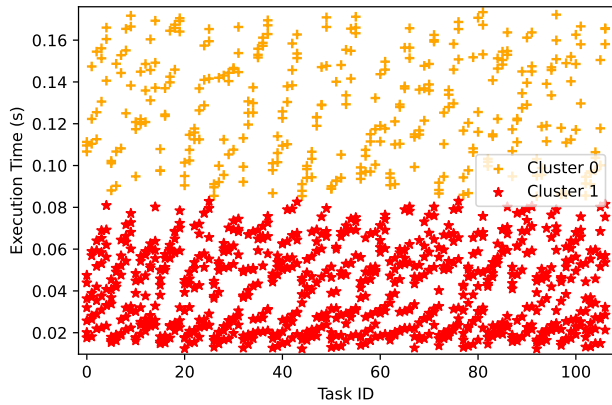


Fig. 7. Data processing execution times on 20 nodes with 8 individual cores.

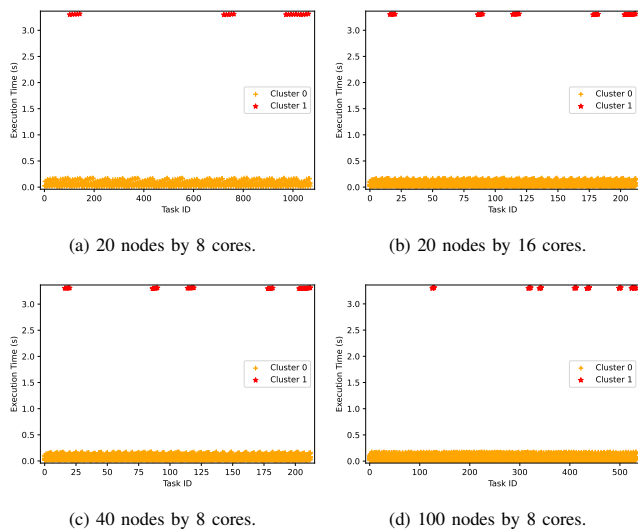


Fig. 8. K-means clustering of run times from experimental scenarios.

score is higher than a threshold lower-bound value  $Z_x$  but less than a threshold upper bound value  $Z_y$ ; backup tasks are required for the cluster with the straggler tasks as seen in line 3. Otherwise, no intervention is applied to the task executions. The remainder of the tasks are then processed on their original nodes as seen in line 6.

Our silhouette score threshold (i.e., lower and upper bound) values were determined from several clustering experiments carried out on our dataset to ensure that the range given satisfies all possible scenarios.

#### IV. EVALUATION

The goal of our evaluation was to assess our approach via two major experiments to prove its applicability. They are (i) strategy implementation overheads experiments (ii) job performance experiments. The experiments enabled us to draw the necessary conclusions on the benefits of using our approach. The experiments aimed to detect and process straggler tasks as backup tasks to improve job performance. The experimental setup is described below.

#### A. Experimental Setup

The following objectives were considered in order to achieve the goal of the experiment:

- To determine the start-time of task processing.
- To determine the completion-time of task processing.
- To capture snapshots of task execution times.
- To capture task run times at specific intervals.
- To terminate straggler tasks.
- To restart straggler tasks on available nodes.

The first two bullet points foster the determination of the overheads introduced by this approach. The last four bullet point ensures the measurements of the jobs improvement performance. The termination and restart of the straggler tasks reduces the high consumption of system resources. The experiment was conducted on our extension of HDMSG MapReduce (a MapReduce simulator with Simgrid as the main backbone) available on GitHub<sup>1</sup>.

In order to utilise HDMSG for the development of this approach, a couple of features had to be added to the simulator to make it applicable. Several methods and classes were created for specific functions. A task monitoring method was created to monitor the tasks running on nodes on the MapReduce Hadoop cluster. This method was responsible for terminating long running tasks. A task rescheduling method was created to move the terminated task to available nodes to be processed as backup tasks. A snapshot capturing method was created to capture the start times and completion times of tasks on nodes. These captured task run times were saved on text files for k-means clustering. A disruption injection method was created to send extra tasks unto arbitrary nodes to serve as background activities. These extra tasks caused the map or reduce task on those nodes to experience longer run times. The methods for the creation of map and reduce tasks were extended to foster the scalability of the framework. A node scheduling class was created to foster the chronological processing of data nodes to enable the capturing of snapshots. Also the class fosters the selection of available nodes as exhibited in the Hadoop infrastructure.

Tasks were divided into ten-equal-length subtasks to simulate the snapshot capturing behaviour with simgrid. This was done to ensure that the snapshot could capture the start-times and completion-times of subtasks. The experiments required the capturing of task processing timelines (i.e. when a particular subtask ends and when the other begins). Therefore, dividing a task into ten-equal-length allowed the runtime behaviours of each subtask to be monitored and captured as snapshots.

In setting up the experiments, the infrastructure of HDMSG with Simgrid were defined. The infrastructure was defined in terms of the following: the number of nodes, CPU cores, bandwidth, latency metrics, and the nodes' speed. Additionally, the number of mappers and reducers, file input size (in megabytes), and block size (HDFS chunk size in megabytes) were configured to foster MR computations.

<sup>1</sup>[https://github.com/EbenezerKomlaGavua/MapReduce\\_Snapshots](https://github.com/EbenezerKomlaGavua/MapReduce_Snapshots)

To determine real life MR cluster infrastructure and application configurations, two surveys about Hadoop cluster requirements were carried out. The first survey focused on identifying typical hadoop configurations and the second one focussed on organisations actively utilising Hadoop clusters for their data processing in industry.

Several keywords such as hadoop clusters (requirements), industry cluster infrastructure (setup, configurations) were employed on several search engines to locate current MRH cluster configurations. The first survey identified Hadoop cluster configurations such as basic or standard deployments, advances deployments, hadoop cluster hardware recommendations for batch processing, in-memory processing, medium data size and large data size. The first survey found that the most used CPU speed was 2-2.5Ghz, data block sizes were between 128-256MB, network bandwidth was 1-10Gbps, cluster nodes was 4-40, number of mappers and reducers were 5-12 per node, disk capacity range was 32 GB to 1.2TB and total system memory was 16-512 GB. All hadoop cluster configurations modes were fully distributed.

The second survey found over one hundred and twenty top companies actively utilising hadoop clusters from several websites. Notable companies amongst the list include Alibaba, AOL, Yahoo, Spotify, Last.fm, Ebay, University of Glasgow-Terrier Team and Criteo. From this list, the modal CPU cores per node identified was eight and the modal cluster nodes was forty.

The findings of the survey fostered the selection of four infrastructure scenarios (displayed in Table I) for our experiments. The experimental scenarios comprise data nodes that ranges from 20 to 100 nodes. The range of CPU cores was 8 to 16. Aside the values displayed in the infrastructure scenarios table, network bandwidth of 10Gbps was simulated for all infrastructure scenarios. The smallest data block size employed was 128MBs. All the experiments were run on fully distributed hadoop cluster mode to foster conformance with industry standard.

The details identified from the survey ensure the modelling of real life applications on the above infrastructures as described below. The number of mappers per node was obtained via eq. (2), as stipulated in<sup>2</sup>:

$$Y = \frac{3Cores}{2}, \quad (2)$$

where  $Y$  is a positive rational number that represents the number of mappers per node.  $Cores$  is a positive integer which represents the CPU cores per node.

The number of reducers per node was obtained via eq. (3), as stipulated in<sup>3</sup>:

$$R = 0.95 \times N \times T, \quad (3)$$

Where  $R$  is a positive rational number representing the number of reducers per node,  $T$  is a positive rational number

TABLE I. EXPERIMENTAL INFRASTRUCTURE SCENARIOS

Features	20N × 8c	20N × 16c	40N × 8c	100N × 8c
No. of Nodes	20	20	40	100
No. of Cores	8	16	8	8
Mappers per node	5	5	11	5
Total Reducers	38	38	76	190
Total Mappers	107	213	213	533
Input Size	13696	27264	27264	68224

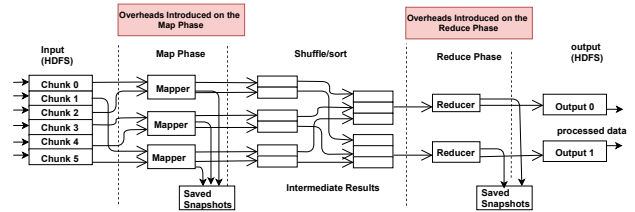


Fig. 9. Structure of algorithm implementation with expected overheads.

representing the mapred tasktracker reduce tasks maximum value.  $T$  is the maximum number of reduce tasks that will be run simultaneously by a task tracker (2 was used, since it is the default maximum value).  $N$  is a positive integer representing the number of nodes running on the cluster.

The proposed approach applies to both mappers and reducers. However, the evaluation was centred on mappers; since the number of mappers are bigger than reducers. Hence, the effects of our approach are expected to be more on mappers than reducers.

### B. Determining the Overheads of our Strategy

This experiment determined the overheads introduced into the infrastructure by the implementation of this approach. The overheads were caused by the effects of the snapshots capturing process on the infrastructure as seen in Fig. 9. The comprehension of the effects of the overheads fosters the appreciation of the challenges and benefits in applying this approach on MRH.

The experiment was conducted on the four data centre scenarios discussed in sub-section IV-A. Mapper tasks with execution times from 0.5 to 2000 seconds were utilized. The range for the experiment was derived via the multiplication of the single values of one, two and five with the power series of ten. The value of negative one produced 0.5 seconds and we scaled the task run times until the graph converged at 2000 seconds. This process was done in order to obtain a scalable range of task run times.

Since this approach involves capturing snapshots during the processing of subtasks, two measurements were taken. These are (i) the commencement of tasks processing and (ii) The completion of task processing. To determine the overhead on a single mapper, the differences between the *completion-times* of the processed portion of the task and the *start-times* of the next portion of that same task are determined (i.e. the period for snapshot capturing). The summation of the differences of these values (i.e. differences between *completion-times* and *start-times*) is subtracted from the task's run times (which is the ten-equal-length subtasks) as shown in Fig. 10. The value realized

<sup>2</sup><https://data-flair.training/forums/topic/how-one-can-decide-for-a-job-how-many-mapper-reducers-are-required/>

<sup>3</sup><https://hadoop.apache.org/docs/r1.2.1/api/org/apache/hadoop/mapred/JobConf.html>

TABLE II. STRATEGY IMPLEMENTATION OVERHEADS

Map time (ms)	Scenario overheads (%)			
	20N×8c	20N×16c	40N×8c	100N×8c
0.5	96	66	66	51
1	53	53	53	39
2	45	36	36	23
5	28	29	29	16
10	24	20	20	12
20	18	15	15	8
50	8	7	7	3
100	7	5	5	1
200	5	3	3	~ 0
500	3	1	1	~ 0
1000	1	~ 0	~ 0	~ 0
2000	~ 0	~ 0	~ 0	~ 0

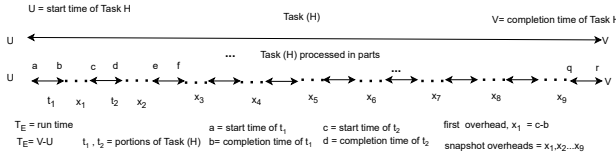


Fig. 10. Strategy overheads.

is the overhead on a single mapper as shown in eq. (4) and Fig. 10.

$$Q^H = 100 - \frac{100(T_E^H - (\sum_{i=1}^{n-1} t_c^{H,i} - t_s^{H,i}))}{T_E^H} \quad (4)$$

Where  $Q^H$  is the overhead of applying our approach on a mapper  $H$  in percentage.  $T_E^H$  is the task runtime of the given mapper.  $t_s^{H,i}$  is the time the  $i^{th}$  snapshot of the mapper  $H$  was started to be captured. Similarly,  $t_c^{H,i}$  is the time when we finished capturing the snapshot of the same task. Equation 4 is exemplified in Fig. 10.  $T_E^H$  is obtained from subtracting  $U$  from  $V$ . The letters  $a$  and  $c$  are the start times of the first two subtasks, whilst  $b$  and  $d$  are the completion times of the first two subtasks. Hence subtracting  $b$  from  $c$  produces the first gap ( $x_1$ ) introduced because of our approach. These gaps  $x_1$  to  $x_n$  are summed up and divided by the task run times to generate the overhead of a task.

### C. Discussion of the Overheads of our Strategy Experiment Results

The overheads of a single mapper were measured on the four data centre scenarios as seen in Tables II and illustrated in Fig. 11.

- **Scenario 20N×8c:** The impact of applying our approach was gradual. The overheads were high at the initial stages of the experiment. However, the impact of the approach caused the high overheads to reduce gradually with longer runtimes, as seen in Fig. 11. Therefore, in such small infrastructures, our approach is only advisable to use with long run times.
- **Scenarios 20N×16c and 40N×8c:** The two scenarios exhibited similar overhead behaviours during task runs. Therefore, only the 40 nodes by 8 cores set up was shown in Fig. 11. The initial overheads observed

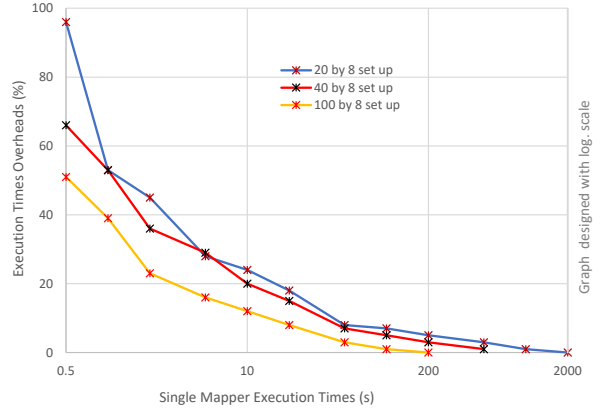


Fig. 11. Cluster overheads.

were 1.5% lower (relatively) than the 20N×8c data centre scenario. The figure shows that the larger infrastructure scenarios converged faster than the previous. Also, this graph shows that the overheads of larger data centres improve better than smaller ones when our approach is applied. Moreover, the overheads of applying our approach with long run times have a higher chance of improving than smaller ones.

- **Scenario 100N×8c:** This demonstrates how our approach deals with larger data-centres. The overhead further reduced over the above scenarios. The initial overheads were 1.9% lower (relatively) than scenario 20N×8c and 1.3% lower (relatively) than the other two scenarios (i.e. 20N×16c and 40N×8c). Also, as the task run times increased, the overheads reduced drastically. Therefore, applying our approach to this scenario shows that initial overheads are mostly lower in large data centres. Additionally, the graph shows that with the large configurations, the overheads reduce faster with long mapper run times than in the other scenarios.

Therefore, from the experiments and industry surveys, it is recommended that infrastructures with 14 to 20 cluster nodes (with eight cores) should use scenario 20N×8c data centre configuration. Infrastructures with 25 to 35 cluster nodes (with eight or sixteen cores) should use our 20N×16c data centre configuration. Infrastructures with 40 to 60 cluster nodes (with eight cores) should use 40N×8c data centre configuration. Finally, infrastructures with 100 to 150 cluster nodes (with eight cores) should use 100N×8c data centre configuration. Aside the above recommendations, our approach is customizable to suit user data configurations preferences.

### D. Application Performance Experiments

This experiment determined the impact of our approach on job performance. Four measurements were taken to evaluate our approach. These are:

- Total execution times when there was *no disruption* on the MapReduce set-up (i.e. a dedicated Hadoop cluster scenario).



- Total execution times when *disruptions* were introduced on arbitrarily nodes on the infrastructure. These disruptions were created to interfere with task processing so that the task will have long run times than expected (this experiment was meant to represent a Hadoop cluster hosted in a multi-tenant environment). These *disruptions* were introduced via the running of extra tasks on arbitrarily nodes which were not linked to the original map or reduce tasks. The extra tasks were designed to consume extra system resources during the map and reduce phase. Also, the *disruptions* represent background services, IO contentions or uneven distribution of resources on data nodes for industry research.
- Total execution times when tasks were terminated and processed as backup tasks (*reschedule*) on a different node. This represents situations where mappers can restore their mid-execution states. This is applied by applications with the capabilities of storing their states during data processing. When such applications get terminated abruptly (due to factors contributing to speculative execution); the applications resume task processing on available nodes from the point they were halted.
- Total execution times when tasks were terminated and processed as backup tasks (*restart*) on a different nodes (providing insight into applications which cannot take advantage of state restoration).

These measurements were utilized to draw the graphs shown in Fig. 12. The graphs of *no disruption*, *reschedule* and *restart* were drawn relative to the *disruption* graphs which are shown in Fig. 13. The graphs were drawn relative to the *disruption* graphs, because we wanted to observe the levels of job improvements in light of the disruptions introduced into the system. The details of the various scenarios are discussed in the next sub-section.

### E. Discussion Performance Experiments Results

First, Fig. 12a displays the behaviour of scenario 20N×8c data centre when our approach was applied. The task improvement on this data centre was gradual as seen in the figure. In relation to the *disruption* graph, the slope began from above 80% and reduced gradually below 10%. Furthermore, *reschedule* backup tasks improve better with our approach than *restart* backup tasks after disruption.

Second, scenario 20N×16c data centre demonstrated considerably more job performance improvement than the previous data centre as most of the graphs were below the 80% mark as seen in Fig. 12b. Also, tasks that transfer their states perform better with our approach than those that cannot. Tasks with long run times exhibited big improvements as their values were below 10% relative to disruption. This means that as the tasks are processed for long run times, the effects of the disruptions were reduced as the graphs approached the 0% mark. For industry practitioners, it is advisable to apply our approach for long run times.

Third, scenario 40N×8c data centre improved more compared to the previous two scenarios as seen in Fig. 12c. Finally,

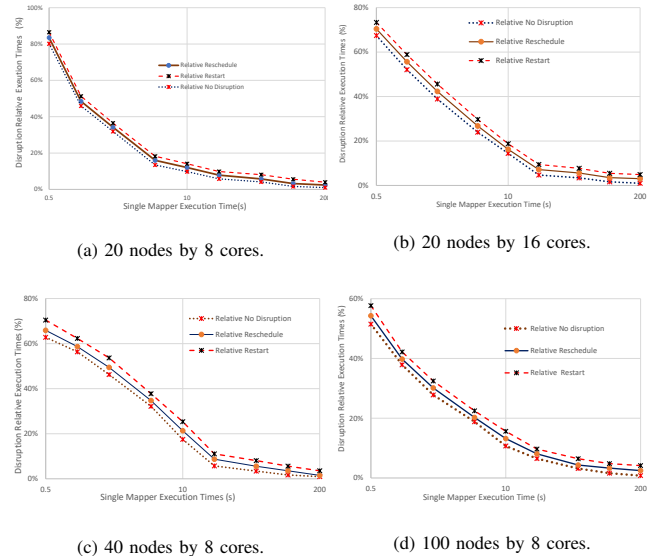


Fig. 12. Tasks improvement experimental scenarios.

TABLE III. KMEANS CLUSTERING SILHOUETTE SCORES

Fig.	Silhouette Scores
7	0.685
8a	0.985
8b	0.985
8c	0.985
8d	0.985

scenario 100N×8c improved more than all the previous scenarios as the graph showed a gradual improvement from below the 60% as seen in Fig. 12d. The figure showed that tasks with long run times had higher chances of improvement in this data centre. As most of the graphs were below 10% relative to disruption. Also, *reschedule* backup tasks improved much better than the (*restart*) backup tasks. Since the *reschedule* backup tasks have the capability to save their states, it was easily for them to continue data processing when they were moved to other nodes. In contrast, the *restart* backup tasks do not store their states, hence they could not reschedule their states, which delayed their task processing durations when moved to other nodes.

In conclusion, larger data centres have a higher chance of improvement when applying this approach. This approach works better with larger data centres because the sizes fosters scalability with long run times, which also ensures reduction in system overheads.

### F. Disruption Identification with K-Means Clustering

The task run times captured during the experiments were utilized for the K-means clustering. Two categories of results were observed after the clustering. Disruption-induced and disruption-free categories. The straggler tasks formed the disruption-biased data clusters are seen in Fig. 8a to 8d on page 17. The large magnitudes of the straggler tasks, enabled k-means to properly create the two categories.

Silhouette score was applied to validate the data clusters

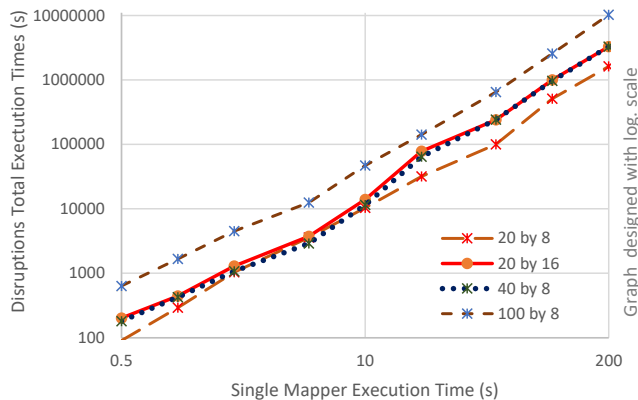


Fig. 13. Scenario disruptions.

and to foster overhead reduction in the application of the kmeans clustering algorithm. Once the dissimilarities between the dataset was detected, the clustering process was stopped. Table III shows the disruption-free and disruption-induced map task runtime data clusters and their respective silhouette scores. The result of Fig. 7 was closer to our silhouette score threshold lower-bound value. Since the data clusters created were not separated enough to ensure task transfer, the task were not shifted to other nodes. However, the rest of the experiments show significantly higher silhouette scores resulting in identifiable straggler tasks. The straggler tasks were then moved to available nodes as either a *reschedule* or *restart* backup tasks.

## V. CONCLUSION

This paper proposed a new speculative execution approach to estimate task run times with consistent global snapshots and K-Means clustering. Our approach applied two algorithms to monitor and capture task run times as snapshots. A K-means clustering technique was applied to classify the captured run times into two categories (fast and straggler tasks). We applied a silhouette score as a decision-making tool to determine when to process backup tasks on available nodes. The silhouette scores also helped to reduce the number of iterations by the K-means. We evaluated our approach on different data centre configurations. These were selected based on a survey of industry requirements for Hadoop clusters and applications. Our experiments were focused on two objectives; (i) the overheads caused by implementing our approach and (ii) job performance improvements. Our experiments enabled us to show that (i) the overheads caused by applying our approach were reduced faster with large data centres than the smaller data centres. The overheads reduced by 1.9%, 1.5% and 1.3% (comparatively) as the size of the data centre and the task run times increased. (ii) Mapper tasks with typical longer task durations had better chances for improvements. The graphs of the longer mappers were below 10% relative to the disruptions introduced. This showed that the effects of the disruptions were reduced and became more negligible. This approach measured the job performance improvement achieved via the *restart* back up tasks. A further work was done for applications capable of transferring their states as *reschedule* back up tasks. Most of

the previous approaches did not consider measurements for *reschedule* back up tasks.

For future work, we consider implementing auto-scaling algorithms on MapReduce Hadoop clouds. Our Snapshot capturing algorithm will be applied to foster a comparison with the job performance approach. Also, a couple of classification and clustering techniques will be considered to provide further extensions.

## ACKNOWLEDGMENT



This research was supported by the Hungarian Scientific Research Fund under the grant number OTKA FK 131793.

## REFERENCES

- [1] T. White, *Hadoop: The definitive guide*. O'Reilly Media, Inc., 2012.
- [2] D. E. O'Leary, "Artificial intelligence and big data," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 96–99, 2013.
- [3] S. De and M. Panjwani, "A comparative study on distributed file systems," in *Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough*. Springer, 2021, pp. 43–51.
- [4] H. Xu and W. C. Lau, "Speculative execution for a single job in a mapreduce-like system," in *2014 IEEE 7th International Conference on Cloud Computing*, 2014, pp. 586–593.
- [5] M. Zaharia, A. Konwinski, A. D. Joseph, R. H. Katz, and I. Stoica, "Improving mapreduce performance in heterogeneous environments," in *Osd*, vol. 8, no. 4, 2008, p. 7.
- [6] X. Huang, L. Zhang, R. Li, L. Wan, and K. Li, "Novel heuristic speculative execution strategies in heterogeneous distributed environments," *Computers & Electrical Engineering*, vol. 50, pp. 166–179, 2016. [Online]. Available: <https://doi.org/10.1016/j.compeleceng.2015.06.013>
- [7] Q. Chen, D. Zhang, M. Guo, Q. Deng, and S. Guo, "Samr: A self-adaptive mapreduce scheduling algorithm in heterogeneous environment," in *2010 10th IEEE International Conference on Computer and Information Technology*. IEEE, 2010, pp. 2736–2743. Available: <https://doi.org/10.1109/cit.2010.458>
- [8] X. Sun, C. He, and Y. Lu, "Esamr: An enhanced self-adaptive mapreduce scheduling algorithm," in *2012 IEEE 18th International Conference on Parallel and Distributed Systems*. IEEE, 2012, pp. 148–155. [Online]. Available: <https://doi.org/10.1109/icpads.2012.30>
- [9] Y. Li, Q. Yang, S. Lai, and B. Li, "A new speculative execution algorithm based on c4. 5 decision tree for hadoop," in *International Conference of Young Computer Scientists, Engineers and Educators*. Springer, 2015, pp. 284–291.
- [10] H. B. Abdalla, A. M. Ahmed, and M. A. Al Sibaheec, "Optimization driven mapreduce framework for indexing and retrieval of big data," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 14, no. 5, pp. 1886–1908, 2020. [Online]. Available: <https://doi.org/10.3837/tiis.2020.05.002>
- [11] I. A. T. Hashem, N. B. Anuar, M. Marjani, E. Ahmed, H. Chiroma, A. Firdaus, M. T. Abdullah, F. Alotaibi, W. K. M. Ali, I. Yaqoob *et al.*, "Mapreduce scheduling algorithms: a review," *The Journal of Supercomputing*, vol. 76, no. 7, pp. 4915–4945, 2020.
- [12] K. Kalia and N. Gupta, "Analysis of hadoop mapreduce scheduling in heterogeneous environment," *Ain Shams Engineering Journal*, vol. 12, no. 1, pp. 1101–1110, 2021. Available: <https://doi.org/10.1016/j.asej.2020.06.009>
- [13] M. Sun, H. Zhuang, C. Li, K. Lu, and X. Zhou, "Scheduling algorithm based on prefetching in mapreduce clusters," *Applied Soft Computing*, vol. 38, pp. 1109–1118, 2016. [Online]. Available: <https://doi.org/10.1016/j.asoc.2015.04.039>
- [14] G. Ananthanarayanan, A. Ghodsi, S. Shenker, and I. Stoica, "Effective straggler mitigation: Attack of the clones," in *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*, 2013, pp. 185–198.

- [15] S. Sakr, A. Liu, and A. G. Fayoumi, "The family of mapreduce and large-scale data processing systems," *ACM Computing Surveys (CSUR)*, vol. 46, no. 1, pp. 1–44, 2013. [Online]. Available: <https://doi.org/10.1201/b171112-6>
- [16] Q. Chen, C. Liu, and Z. Xiao, "Improving mapreduce performance using smart speculative execution strategy," *IEEE Transactions on Computers*, vol. 63, no. 4, pp. 954–967, 2013. [Online]. Available: <https://doi.org/10.1109/tc.2013.15>
- [17] A. Cornuéjols, C. Wemmert, P. Gañçarski, and Y. Bennani, "Collaborative clustering: Why, when, what and how," *Information Fusion*, vol. 39, pp. 81–95, 2018. [Online]. Available: <https://doi.org/10.1016/j.inffus.2017.04.008>
- [18] D. J. Bora, D. Gupta, and A. Kumar, "A comparative study between fuzzy clustering algorithm and hard clustering algorithm," *arXiv preprint arXiv:1404.6059*, 2014. [Online]. Available: <https://doi.org/10.3390/j2020016>
- [19] C. Yuan and H. Yang, "Research on k-value selection method of k-means clustering algorithm," *J*, vol. 2, no. 2, pp. 226–235, 2019. [Online]. Available: <https://doi.org/10.14445/22312803/ijctt-v10p119>
- [20] T. Gupta and S. P. Panda, "Clustering validation of clara and k-means using silhouette & dunn measures on iris dataset," in *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*. IEEE, 2019, pp. 10–13. [Online]. Available: <https://doi.org/10.1109/comitcon.2019.8862199>
- [21] R. Ünlü and P. Xanthopoulos, "Estimating the number of clusters in a dataset via consensus clustering," *Expert Systems with Applications*, vol. 125, pp. 33–39, 2019. [Online]. Available: <https://doi.org/10.1016/j.eswa.2019.01.074>
- [22] H. B. Zhou and J. T. Gao, "Automatic method for determining cluster number based on silhouette coefficient," in *Advanced materials research*, vol. 951. Trans Tech Publ, 2014, pp. 227–230. [Online]. Available: <https://doi.org/10.4028/www.scientific.net/amr.951.227>

# Recognizing Safe Drinking Water and Predicting Water Quality Index using Machine Learning Framework

Mohamed Torky<sup>1</sup>, Ali Bakhiet<sup>2</sup>, Mohamed Bakrey<sup>3</sup>, Ahmed Adel Ismail<sup>4</sup>, Ahmed I. B. EL Seddawy<sup>5</sup>

Faculty of Artificial Intelligence, Egyptian Russian University (ERU), Badr City, Egypt<sup>1</sup>

Higher Institute of Computer Science and Information Systems, Culture & Science City, Giza, Egypt<sup>2,3</sup>

The Higher Institute of Computer and Information Systems, Abo Qir Alexandria 21913, Egypt<sup>4</sup>

Arab Academy for Science and Technology and Maritime Transport, Cairo, Egypt<sup>5</sup>

**Abstract**—Water quality monitoring, analysis, and prediction have emerged as important challenges in several uses of water in our life. Recent water quality problems have raised the need for artificial intelligence (AI) models for analyzing water quality, classifying water samples, and predicting water quality index (WQI). In this paper, a machine-learning framework has been proposed for classify drinking water samples (safe/unsafe) and predicting water quality index. The classification tier of the proposed framework consists of nine machine-learning models, which have been applied, tested, validated, and compared for classifying drinking water samples into two classes (safe/unsafe) based on a benchmark dataset. The regression tier consists of six regression models that have been applied to the same dataset for predicting WQI. The experimental results clarified good classification results for the nine models with average accuracy, of 94.7%. However, the obtained results showed the superiority of Random Forest (RF), and Light Gradient Boosting Machine (Light GBM) models in recognizing safe drinking water samples regarding training and testing accuracy compared to the other models in the proposed framework. Moreover, the regression analysis results proved the superiority of LGBM regression, and Extra Trees Regression models in predicting WQI according to training, testing accuracy, 0.99%, and 0.95%, respectively. Moreover, the mean absolute error (MAE) results proved that the same models achieved less error rate, 10% than other applied regression models. These findings have significant implications for the understanding of how novel deep learning models can be developed for predicting water quality, which is suitable for other environmental and industrial purposes.

**Keywords**—Water quality; artificial intelligence; machine learning; deep learning; classification analysis; and regression analysis

## I. INTRODUCTION

In the new green economy, monitoring and evaluating water quality is a central issue for the life of all organisms. Using the classical monitoring ways that depend on chemical monitoring is not enough to evaluate the consequences of some influences and stresses, as predicting the interactive effects of different chemical variables on water microorganisms is very difficult [1]. Rapid industrial development has deteriorated water quality at an alarming rate. In addition, the infrastructure, with the absence of public awareness, and the low quality of hygiene, greatly affects the quality of drinking water [2].

Polluted drinking water is very serious and can adversely affect organisms' health, as well as many environmental, and infrastructural impacts. According to a United Nations (UN) report, roughly, more than 1.5 million people die every year due to water-polluted diseases. In third-world countries, it has been declared that 80% of health issues are due to polluted water. Moreover, 2.5 billion illnesses and five million deaths are reported annually [3], and these are truly terrifying numbers.

Due to the lack of robust water monitoring techniques, many countries are unable to enhance their water systems and there are shortcomings to produce effective water recovery systems. These shortcomings may lead to a greater level of uncertainty when developing water resource management policies [4].

Recently, there has been a marked increase in the development of rapidly developing biological monitoring and biological assessment tools for water resources that are reliable enough to manage many degraded water bodies in the USA, Europe, South Africa, and Australia [5]. However, with the huge increase in data generated by monitoring devices and the futility of manual coding, the shortcomings began to appear in those systems due to the lack of an effective mechanism for processing that huge data. However, with the growth of artificial intelligence based on machine learning and deep learning techniques, it can introduce a perfect solution to that problem, such as artificial intelligence is characterized by many predictions, clustering, and classification techniques to produce effective solutions to water quality problems [6]. Research of the past decades has focused largely on analyzing the water quality of rivers based on artificial intelligence (AI) techniques [7]. Using AI models, water quality forecasting, classification, and risk assessment can be achieved easily. Moreover, advanced early warning systems and effective management policies can be designed to add more control and monitoring services to rivers and water bodies [8, 9].

In this paper, a proposed machine learning framework has been introduced for analyzing water quality. It consists of two subsystems; the first subsystem is responsible for classifying water quality based on nine AI models that have been applied, tested, and compared to classify various samples of drinking water as safe to drink or unsafe to drink. The applied nine AI

models are: Extreme Gradient Boosting (XGBoost) [10], Light Gradient Boosting Machine (Light GBM) [11], Decision Tree (DT) [12], Extra Tree (ET) [13], Multi-layer Perceptron (MLP) [14], Gradient Boosting (GB) [15], Support Vector Machine (SVM) [16], Artificial Neural Network (ANN) classification [17], and Random Forest (RF) Classifier [18]. The second subsystem is responsible for predicting water quality index (WQI) based on six regression models, LGBM regression, XGB regression, ExtraTrees regression, DT Regression, RF regression, and linear regression. These models have been applied to a dataset called Water quality, which was downloaded from [19]. The experimental results proved the superiority of the LightGBM model compared with the other eight AI models with an accuracy of 97% in classifying water samples to recognize the safe drinking water samples. Moreover, the predictive analysis of the used regression models clarified outperforms of LGBM regression, and Extra Trees Regression models in predicting water quality index according to training accuracy, testing accuracy, and mean absolute error (MAE) compared to the other four regression models.

The rest of this article is designed as follows: Section II reviews the related work. Section III explains the proposed machine-learning framework for analyzing water quality. Section IV presents and discusses the implementation results. Section V presents the conclusion of this work.

## II. LITERATURE REVIEW

A growing body of literature has investigated the efficiency of using machine and deep learning models for monitoring, analyzing, and predicting water quality index. The literature introduced some reviews that discuss various AI models for solving water quality prediction problems [9,20,21]. There are several large cross-sectional studies, which introduces multiple machine and deep learning to predict water quality index.

Ali Najah et al. [22] applied four machine learning models, an enhanced Wavelet De-noising Techniques (WDT)-based Neuro-Fuzzy Inference System (WDT-ANFIS), Adaptive Radial Basis Function Neural Networks (RBF-ANN), Neuro-Fuzzy Inference System (ANFIS), Multi-Layer Perceptron Neural Networks (MLP-ANN), and to predict water quality parameters (i.e. pH, ammonia nitrogen (AN), and suspended solids (SS)) of Johor River in Malaysia. The experimental results clarified outperform of the WDT-ANFIS model in prediction accuracy for all the water quality parameters compared to the other three used models.

Amir Hamzeh et al. [23] used the support vector machine (SVM) algorithm, Artificial Neural Network (ANN), and group method of data handling (GMDH) models for analyzing the water quality prediction of Tireh River in Iran. Different types of the kernel and transfer functions were validated and tested, and the practical results clarified that both ANN and SVM are better models than GMDH in predicting the water quality of Tireh River.

Umair Ahmed et al [24] introduced supervised learning models for evaluating WQI prediction based on four features of water elements, namely, turbidity, temperature, pH, and total dissolved solids. The proposed models achieved acceptable

accuracy and fewer error rates using a minimal number of features in predicting the WQI in real-time.

Abubakr Saeed et al. [25] proposed an efficient machine learning algorithm based on the SVM model to forecast the WQI of Langat River Basin based on the investigation of six variables (Dissolved Oxygen (DO), pH, Chemical Oxygen Demand (COD), Suspended Solids (SS), Ammonia Nitrogen (AN), and Biochemical Oxygen Demand (BOD)) of dual reservoirs that are located in the catchment. The experimental results showed that this model could accurately predict WQI value with small mean absolute error.

Mourad Azrou et al. [26] investigated the efficiency of machine learning algorithms for evaluating WQI prediction value based on four water features: pH, temperature, turbidity, and coliforms. The experimental results have proven the efficiency of used regression algorithms in predicting WQI. Moreover, the artificial neural network proved that it is the most highly efficient model in classifying water quality compared to other models in the literature.

They H et al. [27] utilized advanced AI models to evaluate WQI prediction value and classifying water goodness. The authors applied nonlinear autoregressive neural networks (NARNET) and long short-term memory (LSTM) as deep learning algorithms for predicting WQI. Moreover, three learning techniques, namely, K-nearest neighbor (K-NN), Naive Bayes, and SVM have been applied for the water quality classification task. The Prediction results showed that the NARNET algorithm performed slightly better than the LSTM for predicting WQI values. On the other hand, the SVM model has achieved the greatest accuracy (97.01%) for water goodness classification compared to the other classification models.

Siti Nur Mahfuzah et al. [28] investigated the efficiency of two machine learning algorithms, the Random Forest algorithm and the Random Tree algorithm for Classifying River Water Quality. The practical results have proven that Random Forest gives a higher classification accuracy compared to the Random Tree algorithm.

Junhao Wu et al. [29] proposed a hybrid model based on discrete wavelet transform (DWT), an ANN model, and LSTM model to predict the water goodness of the Jinjiang River. The prediction results clarified the efficiency of the proposed hybrid model in predicting water quality index compared to other models such as the ARIMA model, the LSTM model, nonlinear autoregression (NAR) model, the ANN-LSTM model, multi-layer perceptron model, and the CNN-LSTM model.

NguyenHien Than et al. [30] investigated water quality monitoring for the Dong Nai River at different times based on a novel architecture of the neural network model FFNN, and LSTM-MA hybrid model at different time series. The validation results proved that The LSTM-MA model provided more reliable prediction and achieved faster training time than the NAR, NAR-MA, ARIMA, and LSTM models. Moreover, the proposed hybrid model produced classification results for water quality in close agreement with the actual monitoring data.

Other hybrid machines and deep learning models have been developed for investigating water quality index, for example, one-dimensional residual CNN (1-DRCNN) and bi-directional gated recurrent units (BiGRU) have been utilized for predicting Water Quality in the Luan River [31]. Moreover, a hybrid deep learning model based on the CNN and LSTM model has been applied, tested, and compared for predicting water goodness based on real-time monitoring of water quality variables [32].

### III. WATER QUALITY ANALYSIS FRAMEWORK

Automatic analyzing drinking water quality from a given dataset, a framework consisting of two phases is proposed. The first phase is responsible for classifying water samples from a given dataset into two classes, safe or unsafe for drinking based on nine classification algorithms, whereas, the second phase is responsible for predicting the water quality index (WQI) based on six regression algorithms. In the following, the two phases are discussed in more detail:

#### A. Phase 1: Water Samples Classifications

To classify water samples to recognize safe drinking water samples, nine-machine learning techniques have been used, tested, and compared. Fig. 1 depicts how these models can be used for classifying water samples from a given dataset. The classification phase starts by doing a preprocessing step for cleaning, splitting, and resampling the used dataset. In the second step, the given dataset is divided into training (70%) and testing (30%) data parts. The third step focuses on extracting water features that may impact water quality through a feature selection step. The final step, the classification step sequentially calls nine classification algorithms (i.e. learning model) one after one for performing the classification task. The used classification models can be briefly described as follows:

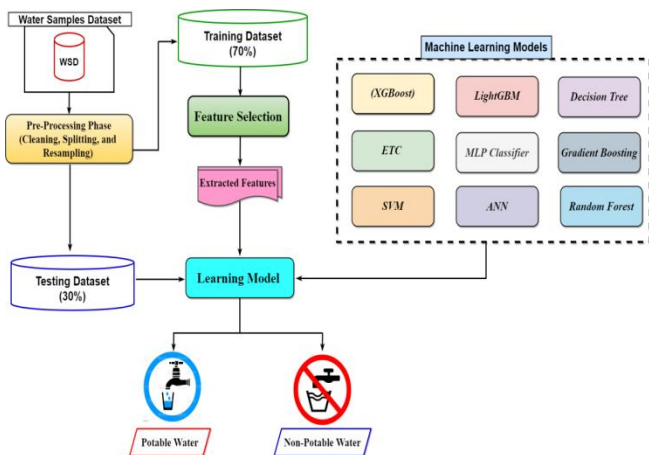


Fig. 1. Water quality classification model.

1) *Extreme Gradient Boosting (XGBoost)*: It is depending on supervised machine learning, decision trees, ensemble learning, and gradient boosting. It is one of the most powerful techniques for building stochastic models for regression, classification, and ranking problems [33]. It provides a parallel tree boosting approach to fix errors made by prior boosted tree models [34].

2) *Light Gradient Boosting Machine (Light GBM)*: It has been developed by Microsoft, which is a popular algorithm

used for ranking and classification problems. Its structure is also based on decision tree models. LightGBM is being distinguished by training speed and accurate prediction results. This is because of adding an automatic feature selection procedure as well as focusing on boosting instances with greater gradients [35].

3) *Decision Tree (DT)*: It is a common supervised learning algorithm used for regression and classification problems [12]. The idea is to use learning decision rules deduced from the data features to perform classification or prediction tasks. What makes DT an effective classification model is: 1) the DT model can be prepared with little data. 2) Training a DT model is logarithmic in the number of data points. 3) A DT model can be validated by statistical tests. 4) Its performance doesn't affect any violation in predefined assumptions with the original model from which the data were created. 5) DT models can be visualized easily and can be understood without mysterious [36].

4) *Extra Tree Classifier (ETC)*: It is a class of ensemble learning approaches. The classification results are collected from a forest of several de-correlated DT models [37]. It differs from Random Forest Classifier in DT constructions way, where DT models are constructed in a "forest". The forest construction and creation of multiple de-correlated DT models of this classifier are based on extracting a random sample of features that leads to the best classification results based on some mathematical conditions.

5) *Multi-layer Perceptron Classifier (MLP Classifier)*: It is a class of feed-forward neural network models [38]. There may be multiple nonlinear hidden layers between the input and the output layers for mapping input data to output data. This classifier is based on the functionality of the sigmoid activation function for doing the classification task.

6) *Gradient Boosting Classifier (GBC)*: It is a common boosting classifier algorithm [39]. The functionality of gradient boosting works based on training N Trees based on the repeated fixing errors resulting from the predecessors of predictors to form the ensemble of data. The training step of the GBC model is done by training the predictors with the error labels produced by the predecessor of those predictors. The prediction results of each tree model are based on "a shrinking routine".

7) *Support Vector Machine (SVM) Classifier*: it is a supervised learning model used for both regression and classification problems [40]. The main goal of the SVM model is to identify a hyperplane in an N-dimensional space for classifying data items. The kernel of SVM is a procedure that depends on low-dimensional input space and converts it into higher-dimensional space. Therefore, SVM is suitable for non-linear classification problems. SVM has some advantages that make it an efficient classifier such as memory efficiency, effectiveness in high dimensional cases, and possible to customize kernel functions.

8) *Artificial Neural Network Classification (ANN)*: This class of ANN is one of the simplest types of neural networks

[17]. It is also a fed forward algorithm as it passes information in one direction from input neurons through one or more hidden layers to output neurons. The main advantages of using an ANN classifier are the ability to work with incomplete knowledge, storing information on the entire network, having a distributed memory, and having fault tolerance.

9) *Random Forest Classifier (RF)*: It is a non-linear classification technique, which consists of a group of decision trees. [18]. It integrates multiple decision trees to get more accurate predictions. Each decision tree model is used when employed on its own. This algorithm is called random because they choose predictors randomly at a time of training. In addition, it is called a forest since it takes the result of multiple trees to make a decision. The main advantage of Random forests compared to decision trees is the large number of uncorrelated tree models that work as a single unit will always outperform the individual tree models.

**B. Phase 2: Water Quality Index Prediction**

The second phase of the proposed framework is responsible for the predictive analysis of the water quality index. In this phase, we examined the impact of the water quality index (WQI) in predicting water quality using six regression models. This analysis started by calculating WQI for the dataset using a

mathematical model specified in equations 1, 2, 3, and 4 [41]. After that, six regression models have been applied for predicting water quality. These models are LGBM regression, XGB regression, Extra Trees regression, Decision Tree Regression, Random Forest regression, and linear regression [42]. Fig. 2 explains how the six regression models are applied to predict the water quality index.

$$K = \frac{1}{\sum(\frac{1}{S_i})} \tag{1}$$

Where,  $S_i$  is the standard value for each variable of water elements, and  $K$  is a constant.

Then, the weight value  $W_i$  of each element can be calculated as in equation 2.

$$W_i = \frac{k}{S_i} \tag{2}$$

The Quality Impact  $Q_i$  value for each element in the water dataset can be calculated as in equation 3.

$$Q_i = 100 * \left( \frac{\text{observe values } (oi) - \text{initial value } (li)}{\text{standard value } (si) - \text{initial value } (li)} \right) \tag{3}$$

Finally, the water quality index  $WQI$  can be calculated as in equation 4.

$$WQI = \sum_{i=1}^N W_i * Q_i \tag{4}$$

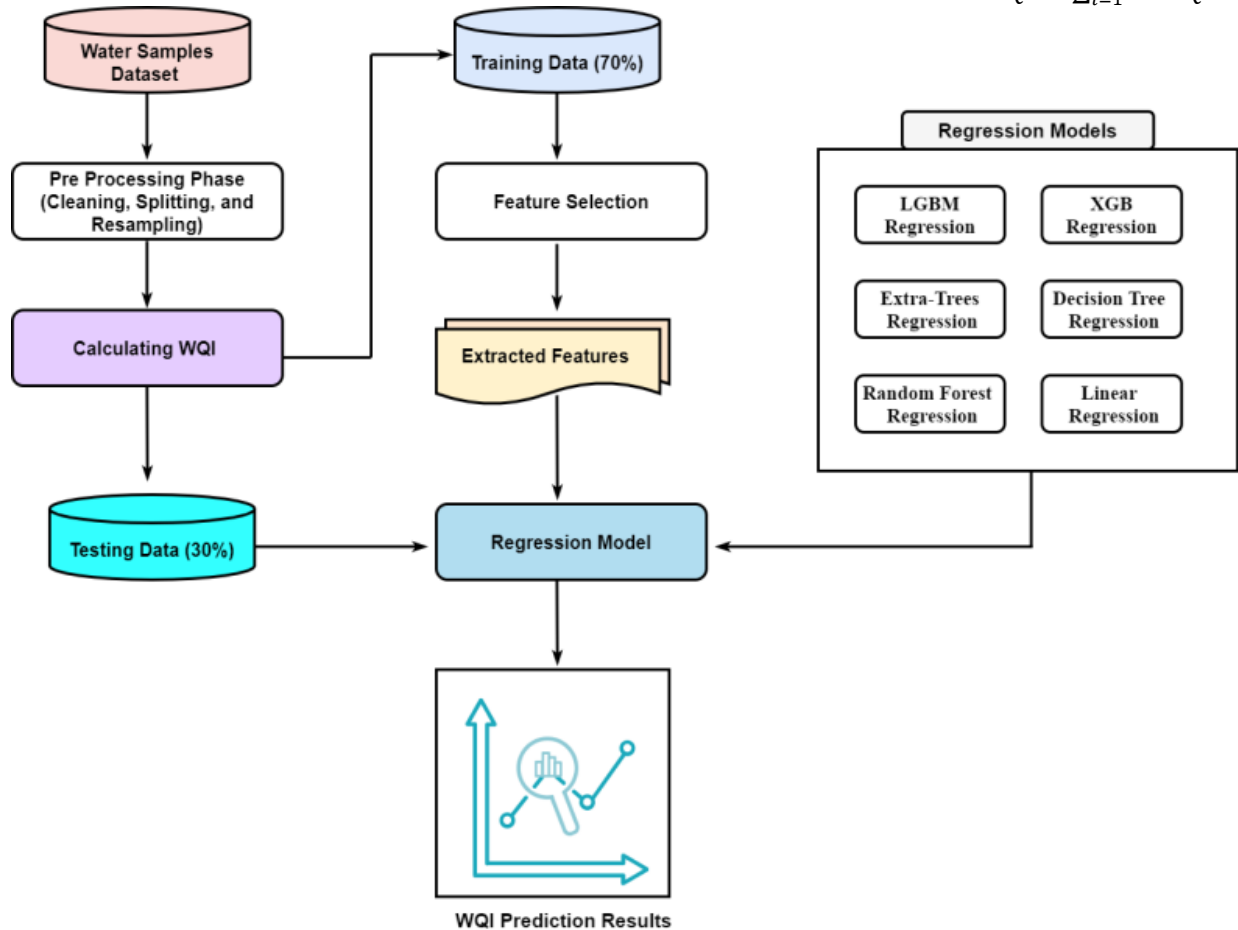


Fig. 2. Water quality index prediction model.

#### IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we present two types of analysis for investigating the efficiency of the proposed machine learning approach in predicting water quality. Subsection A discusses the classification analysis of water samples using nine classifiers, while subsection B discusses the predictive analysis using five regression models:

##### A. Classification Analysis

The first set of analyses examined the efficiency and accuracy of nine machine learning models used in the proposed framework (as explained in section 3.1) for classifying water samples to recognize that good samples are suitable for human drinking. These performances of these models have been applied to a dataset called Water quality, which was downloaded from [19]. The used dataset consists of 7996 samples of water and 19 features (i.e. variables) that impact water quality. The data has been segmented into training data (6396 samples, 19 features), and testing data (1600 samples, 19 features). The main objective was to classify water samples as suitable for human drinking or not suitable for human drinking. The performance of the nine machine learning models used in the proposed framework has been tested and evaluated using twelve measures as detailed in Table I. The best performance among the nine machine learning models according to each measure is being highlighted. The obtained results clarify that although the random forest algorithm achieved the best training accuracy, the Light GBM outperformed the other classifiers in recognizing good water samples regarding testing accuracy, sensitivity, AUC, F1-score, recall, precision, and mean square error. Fig. 3 and 4 present the comparison results of classification analysis metrics and mean square error (MSE) to nine classifiers, respectively. In addition, Fig. 5 to 13 depicts

the performance matrices (or confusion matrices) and the corresponding receiver operating characteristic (ROC) curves of nine machine-learning models, respectively.

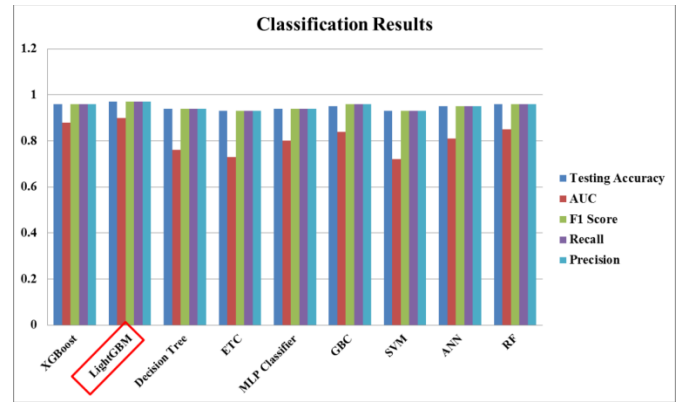


Fig. 3. Comparison results of classification analysis to nine classifiers.

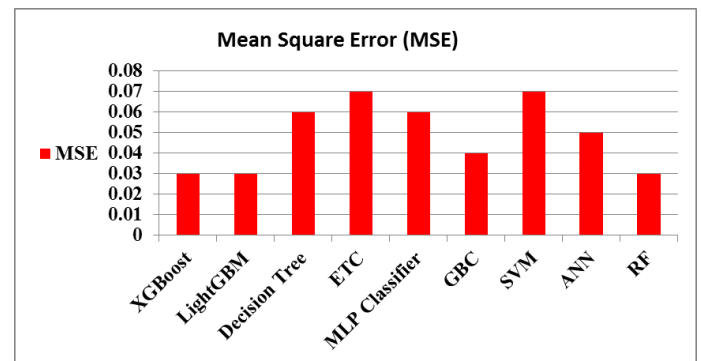
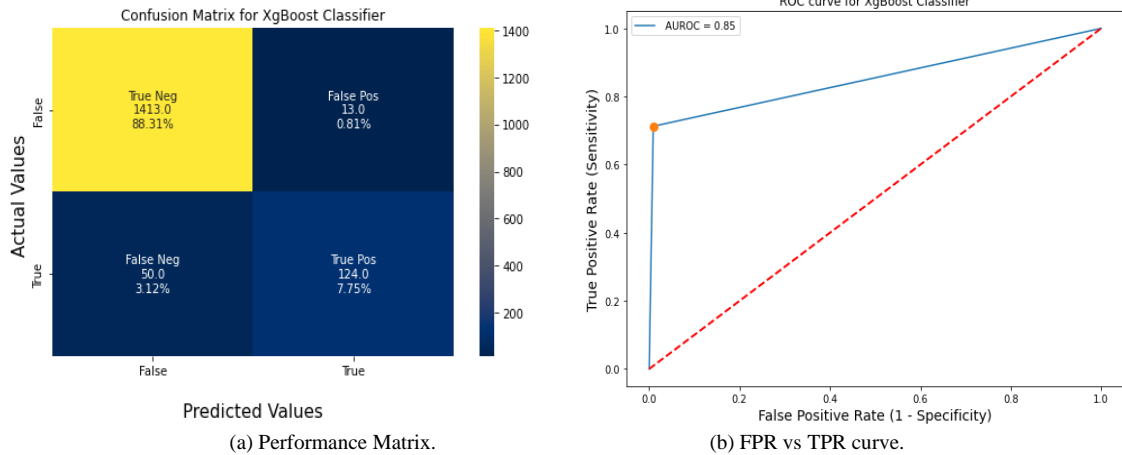


Fig. 4. Comparison of results of Mean Square Error (MSE) to nine classifiers.

TABLE I. PERFORMANCE EVALUATION RESULTS OF NINE MACHINE LEARNING MODELS USED IN THE PROPOSED FRAMEWORK

Measure	XGBoost	LightGBM	Decision Tree	ETC	MLP Classifier	GBC	SVM	ANN	RF
Training accuracy	0.97	0.99	0.94	1.0	0.97	0.97	0.95	0.94	1.0
Testing Accuracy	0.96	0.97	0.94	0.93	0.94	0.95	0.93	0.95	0.96
Sensitivity	0.97	0.98	0.94	0.93	0.95	0.96	0.94	0.96	0.97
Specificity	0.91	0.88	0.88	0.88	0.8	0.91	0.82	0.89	0.94
NPV	0.71	0.81	0.52	0.45	0.62	0.70	0.45	0.63	0.71
AUC	0.88	0.90	0.76	0.73	0.80	0.84	0.72	0.81	0.85
F1 Score	0.96	0.97	0.94	0.93	0.94	0.96	0.93	0.95	0.96
Recall	0.96	0.97	0.94	0.93	0.94	0.96	0.93	0.95	0.96
Precision	0.96	0.97	0.94	0.93	0.94	0.96	0.93	0.95	0.96
Mean SE	0.03	0.03	0.06	0.07	0.06	0.04	0.07	0.05	0.03

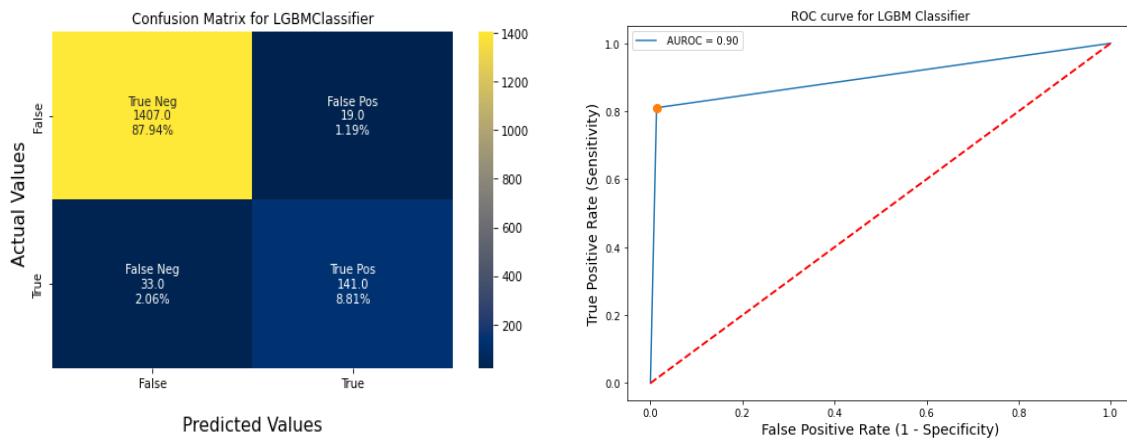




(a) Performance Matrix.

(b) FPR vs TPR curve.

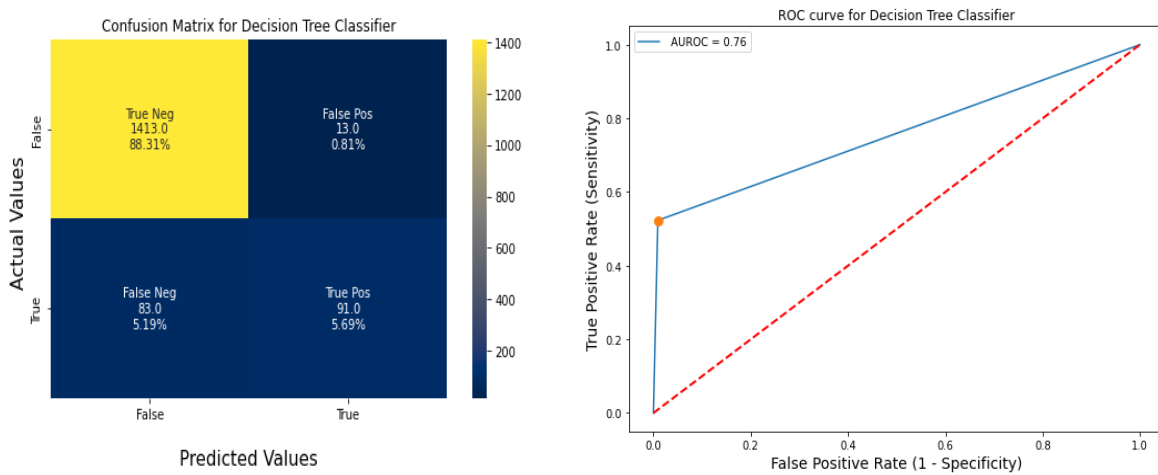
Fig. 5. (a) Performance Matrix and (b) FPR vs TPR curve of XGBoost classifier.



(a) Performance Matrix.

(b) FPR vs TPR curve.

Fig. 6. (a) Performance Matrix and (b) FPR vs TPR curve of LightGBM classifier.



(a) Performance matrix.

(b) FPR vs TPR curve.

Fig. 7. (a) Performance Matrix and (b) FPR vs TPR curve of Decision Tree classifier.

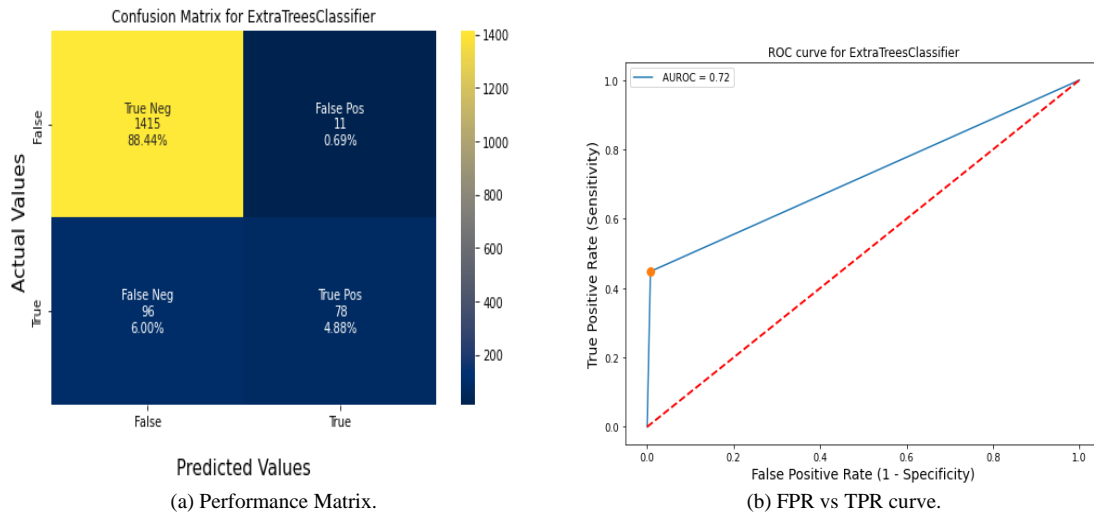


Fig. 8. (a) Performance Matrix and (b) FPR vs TPR curve of Extra Trees Classifier (GBC).

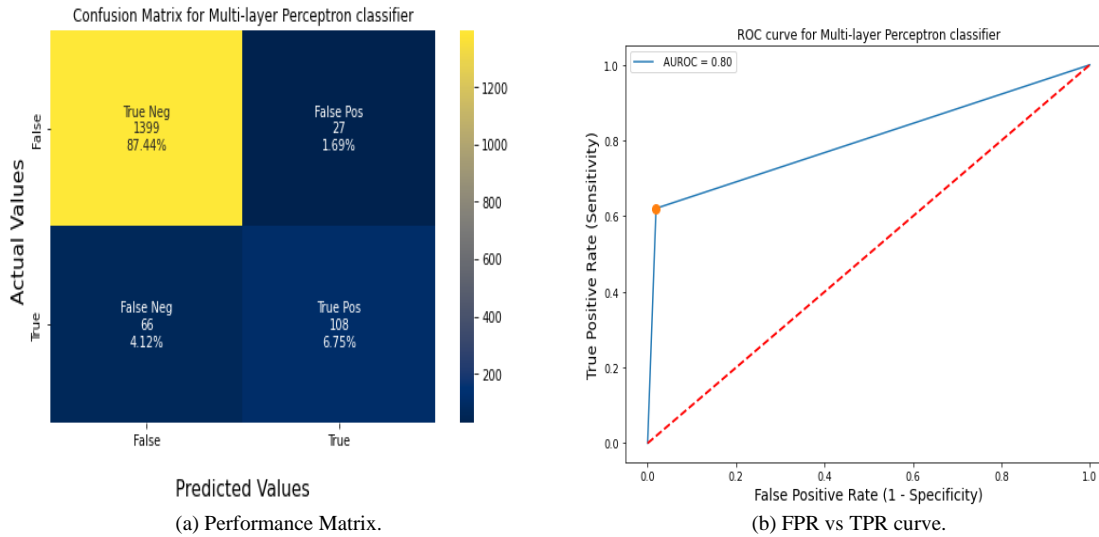


Fig. 9. (a) Performance Matrix and (b) FPR vs TPR curve of MLP Classifier.

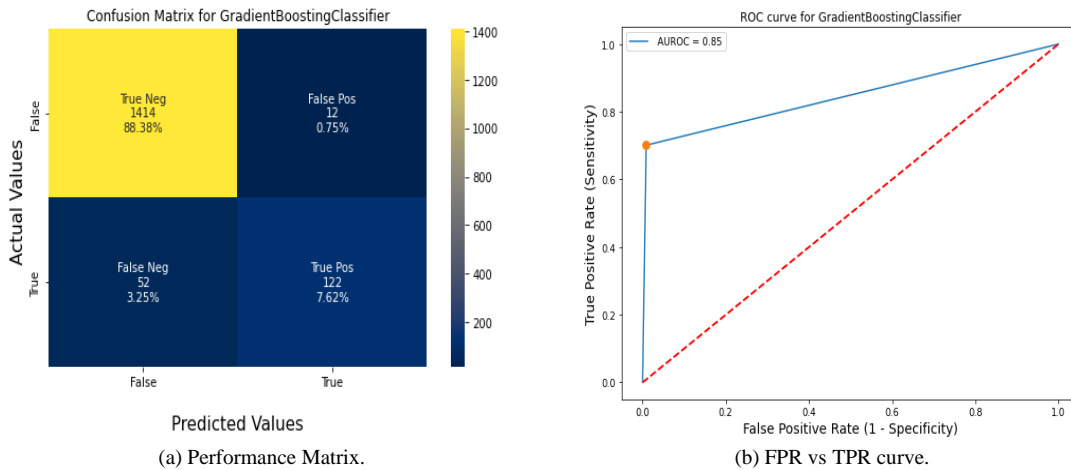
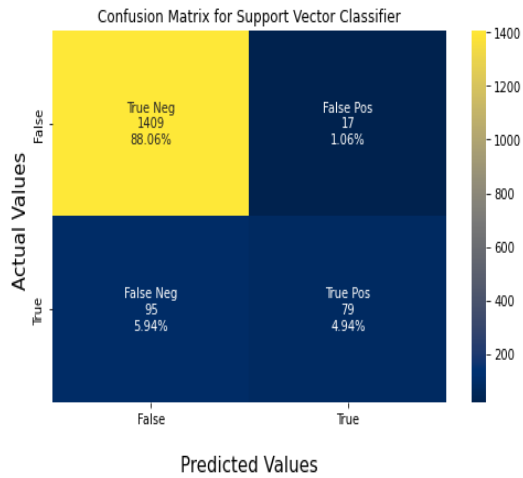
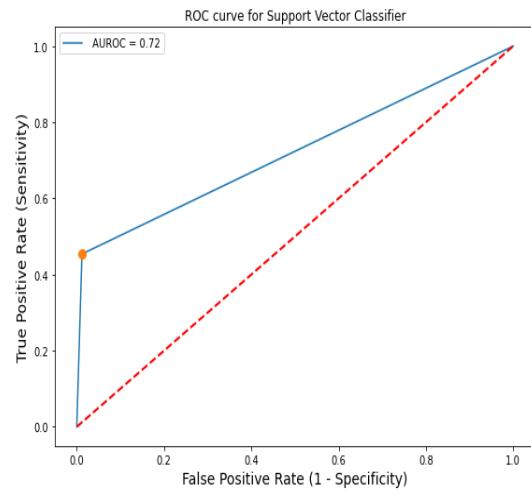


Fig. 10. (a) Performance Matrix and (b) FPR vs TPR curve of GB Classifier.

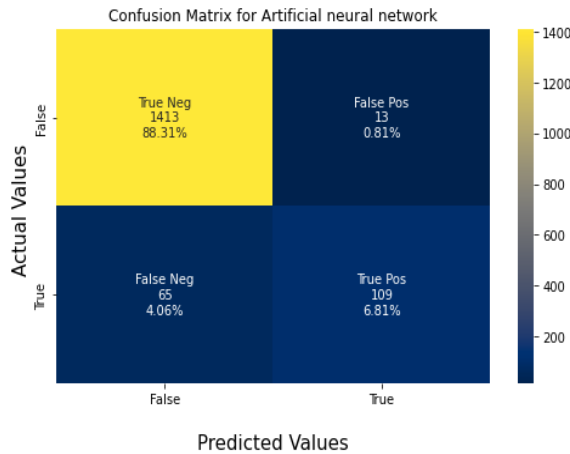


(a) Performance Matrix.

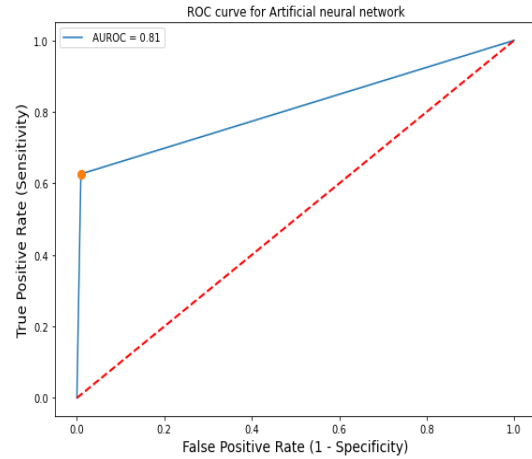


(b) FPR vs TPR curve.

Fig. 11. (a) Performance Matrix and b) FPR vs TPR curve of SVM Classifier.

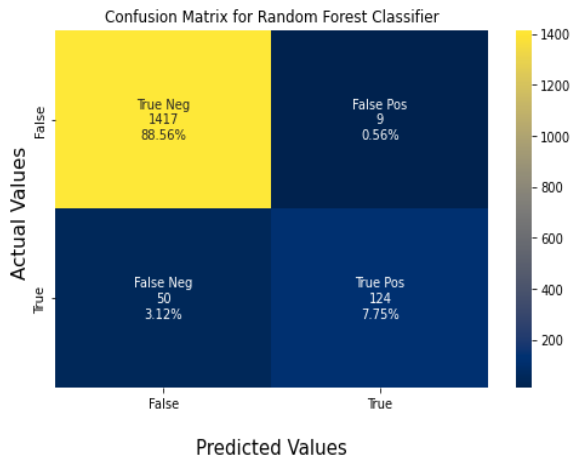


(a) Performance Matrix.

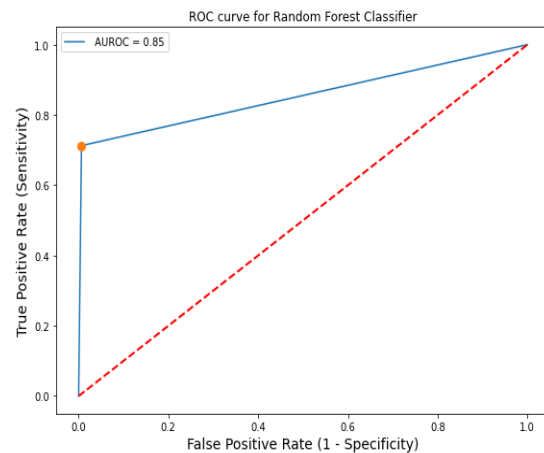


(b) FPR vs TPR curve.

Fig. 12. (a) Performance Matrix and (b) FPR vs TPR curve of ANN Classifier.



(a) Performance Matrix.



(b) FPR vs TPR curve.

Fig. 13. (a) Performance Matrix and (b) FPR vs TPR curve of RF Classifier.

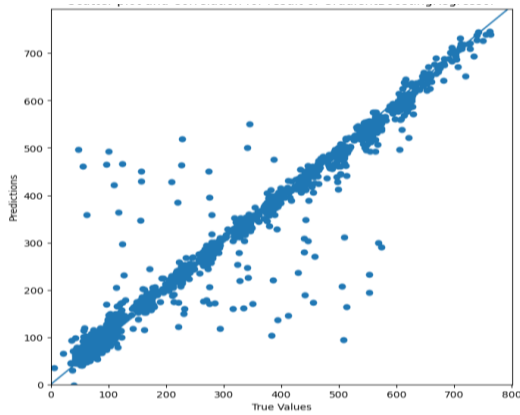
**B. Predictive Analysis**

The second set of analyses examined the efficiency and accuracy of six regression machine learning models used in the proposed framework (as explained in section 3.2) for predicting WQI. Table II summarizes the predictive analysis results of the six regression models after applying the mathematical model of WQI in the dataset. The obtained results have been evaluated based on the common regression metrics, training accuracy, testing accuracy, R2, Adjusted R2, and Mean absolute error (MAE).

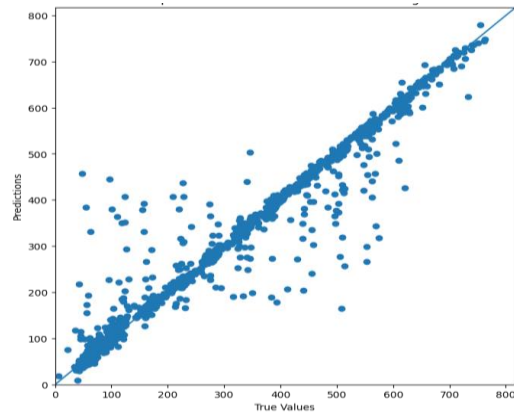
The regression analysis results show the superiority of LGBM regression, and Extra Trees Regression models in predicting water quality index according to training and testing accuracy as well as the mean absolute error (MAE) compared to the other regression models. Fig. 14 to 16 visualizes the prediction results of the six regression models, respectively. Fig. 17 presents the comparison results of regression analysis of the used six regression models.

TABLE II. REGRESSION ANALYSIS RESULTS

Models and Measurement	LGBM Regression	XGB Regression	Extra Trees Regression	Decision Tree Regression	Random Forest Regression	Linear Regression
Training Accuracy	99.0	97.6	99.9	95.4	98.8	92.9
Testing Accuracy	95.5	95.4	95.5	94	94.8	93.5
$R^2$	94.2	95.47	95.55	94.09	90.6	90.6
Adjusted $R^2$	94.1	95.41	95.2	94.02	90.5	90.5
MAE	10.88%	15.754%	10.07%	17.45%	15.35%	19.34%

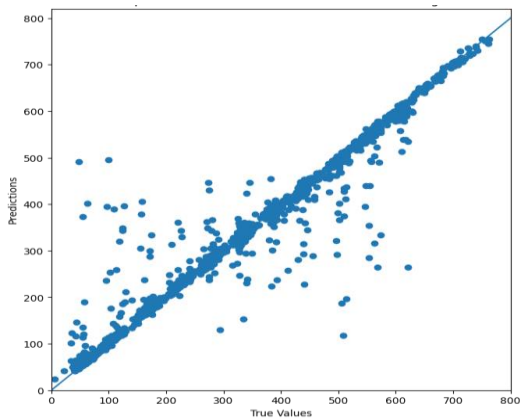


(a) LGBM Regression.

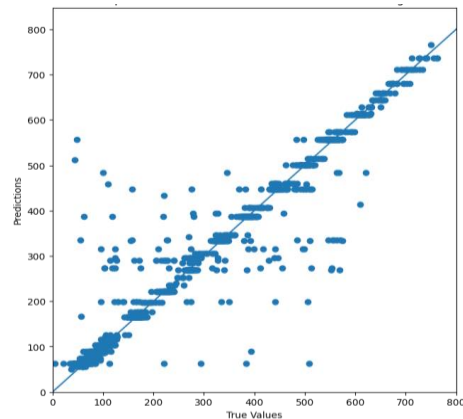


(b) XGB Regression.

Fig. 14. Regression analysis results of a) LGBM Regression and b) XGB Regression.



(a) ET Regression.



(b) DT Regression.

Fig. 15. Regression analysis results of a) ET Regression and b) DT Regression.

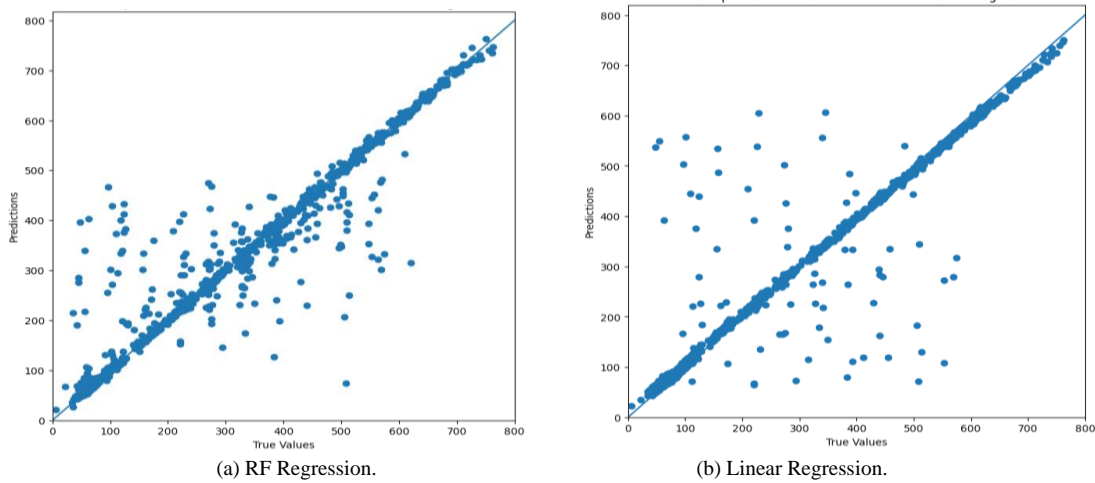


Fig. 16. Regression analysis results of a) RF regression and b) Linear regression.

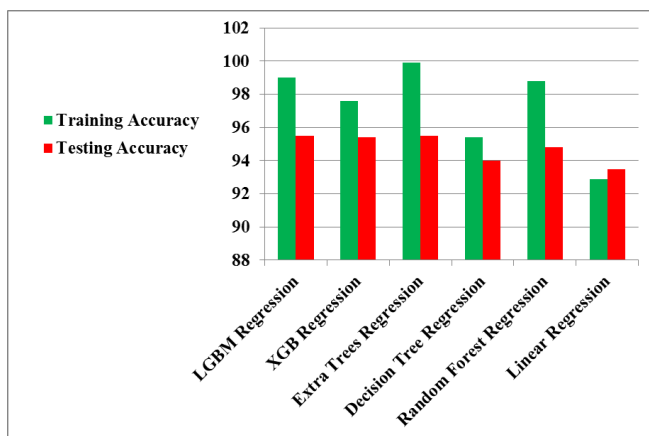


Fig. 17. Comparison results of regression analysis results of the used six regression models.

## V. CONCLUSION

The present article was designed to investigate the efficiency of using a proposed machine-learning framework to classify drinking water samples and predict water quality index. The classification tier of the proposed framework consists of nine classification models, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting (LightGB), Decision Tree (DT), Extra Tree (ET) classifier, Multi-layer Perceptron (MLP) classifier, the Gradient Boosting (GB) classifier, Support Vector Machine (SVM), Artificial Neural Network (ANN), and Random Forest (RF) classifier. The performance of those models has been validated on a benchmark dataset consisting of 7996 water samples, and 19 features. The obtained results clarified good classification results to the nine models with average accuracy. 94.7%. However, the obtained results clarified that, although the Random Forest (RF) algorithm achieved the best training accuracy, 100%, the Light GBM outperformed the other classifiers in recognizing good water samples regarding testing accuracy, 0.97%. The second goal of this study was to investigate the efficiency of the regression tier through applying six regression models for predicting water quality index. The regression analysis clarified the superiority of LGB regression, and Extra Trees Regression models in predicting

water quality index according to training and testing accuracy as well as the mean absolute error (MAE) compared to the other regression models. Taken together, these findings suggest a role for using machine learning models in promoting the analysis and prediction of water quality. Moreover, these results have significant implications for the understanding of how novel deep learning models can be developed for predicting water quality, which is suitable for human drinking, irrigation of plants and crops, and other industrial or environmental purposes.

## REFERENCES

- [1] Wolfram J, Stehle S, Bub S, Petschick LL, Schulz R. Water quality and ecological risks in European surface waters—Monitoring improves while water quality decreases. *Environment International*. 2021 Jul 1;152:106479.
- [2] World Health Organization. A global overview of national regulations and standards for drinking-water quality. [online], available at: <https://apps.who.int/iris/bitstream/handle/10665/350981/9789240023642-eng.pdf?sequence=1> (accessed 19/4/2022).
- [3] United Nations, Water [online], available at: <https://www.un.org/en/global-issues/water> (Accessed 19/4/2022).
- [4] Alamanos A, Mylopoulos N, Loukas A, Gaitanaros D. An integrated multicriteria analysis tool for evaluating water resource management strategies. *Water*. 2018 Dec;10(12):1795.
- [5] Schmutz S, Sendzimir J. *Riverine ecosystem management: Science for governing towards a sustainable future*. Springer Nature; 2018.
- [6] Tung TM, Yaseen ZM. A survey on river water quality modeling using artificial intelligence models: 2000–2020. *Journal of Hydrology*. 2020 Jun 1;585:124670.
- [7] Hmoud Al-Adhaileh M, Waselallah Alsaade F. Modelling and prediction of water quality by using artificial intelligence. *Sustainability*. 2021 Jan;13(8):4259.
- [8] Hameed M, Sharqi SS, Yaseen ZM, Afan HA, Hussain A, Elshafie A. Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in the tropical region, Malaysia. *Neural Computing and Applications*. 2017 Dec;28(1):893-905.
- [9] Chen Y, Song L, Liu Y, Yang L, Li D. A review of the artificial neural network models for water quality prediction. *Applied Sciences*. 2020 Jan;10(17):5776.
- [10] Osman AI, Ahmed AN, Chow MF, Huang YF, El-Shafie A. Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*. 2021 Jun 1;12(2):1545-56.

- [11] Gan M, Pan S, Chen Y, Cheng C, Pan H, Zhu X. Application of the machine learning LightGBM model to the prediction of the water levels of the lower Columbia River. *Journal of Marine Science and Engineering*. 2021 May;9(5):496.
- [12] Lu H, Ma X. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*. 2020 Jun 1;249:126169.
- [13] Upadhyay R, Tanwar PS, Degadwala S. Fracture Type Identification Using Extra Tree Classifier. In 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud)(I-SMAC) 2021 Nov 11 (pp. 1-6). IEEE.
- [14] Haribabu S, Gupta GS, Kumar PN, Rajendran PS. Prediction of Flood by Rainfall Using MLP Classifier of Neural Network Model. In 2021 6th International Conference on Communication and Electronics Systems (ICES) 2021 Jul 8 (pp. 1360-1365). IEEE.
- [15] Khan MS, Islam N, Uddin J, Islam S, Nasir MK. Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. *Journal of King Saud University-Computer and Information Sciences*. 2021 Jun 14.
- [16] Singh KP, Basant N, Gupta S. Support vector machines in water quality management. *Analytics chimica acta*. 2011 Oct 10;703(2):152-62.
- [17] Sulaiman K, Ismail LH, Razi MA, Adnan MS, Ghazali R. Water quality classification using an Artificial Neural Network (ANN). In *IOP Conference Series: Materials Science and Engineering* 2019 Aug 1 (Vol. 601, No. 1, p. 012005). IOP Publishing.
- [18] Ko BC, Kim HH, Nam JY. Classification of potential water bodies using Landsat 8 OLI and a combination of two boosted random forest classifiers. *Sensors*. 2015 Jun;15(6):13763-77.
- [19] Mss.Pants, Water quality, [online], available at <https://www.kaggle.com/datasets/mssmartypants/water-quality>.
- [20] Mustafa HM, Mustapha A, Hayder G, Salisu A. Applications of IoT and Artificial Intelligence in Water Quality Monitoring and Prediction: A Review. In 2020 6th International Conference on Inventive Computation Technologies (ICICT) 2021 Jan 20 (pp. 968-975). IEEE.
- [21] Rajae T, Khani S, Ravansalar M. Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. *Chemometrics and Intelligent Laboratory Systems*. 2020 May 15;200:103978.
- [22] Ahmed AN, Othman FB, Afan HA, Ibrahim RK, Fai CM, Hossain MS, Ehteram M, Elshafie A. Machine learning methods for better water quality prediction. *Journal of Hydrology*. 2019 Nov 1;578:124084.
- [23] Haghiabi AH, Nasrolahi AH, Parsaie A. Water quality prediction using machine learning methods. *Water Quality Research Journal*. 2018 Feb;53(1):3-13.
- [24] Ahmed U, Mumtaz R, Anwar H, Shah AA, Irfan R, García-Nieto J. Efficient water quality prediction using supervised machine learning. *Water*. 2019 Oct 24;11(11):2210.
- [25] Abubakr Yahya AS, Ahmed AN, Binti Othman F, Ibrahim RK, Afan HA, El-Shafie A, Fai CM, Hossain MS, Ehteram M, Elshafie A. Water quality prediction model based support vector machine model for ungauged river catchment under dual scenarios. *Water*. 2019 Jun 13;11(6):1231.
- [26] Azrou M, Mabrouki J, Fattah G, Guezzaz A, Aziz F. Machine learning algorithms for efficient water quality prediction. *Modeling Earth Systems and Environment*. 2022 Jun;8(2):2793-801.
- [27] Aldhyani TH, Al-Yaari M, Alkahtani H, Maashi M. Water quality prediction using artificial intelligence algorithms. *Applied Bionics and Biomechanics*. 2020 Dec 30;2020.
- [28] Nafi SN, Mustapha A, Mostafa SA, Khaleefah SH, Razali MN. Experimenting with two machine learning methods in classifying river water quality. *International Conference on Applied Computing to Support Industry: Innovation and Technology* 2019 Sep 15 (pp. 213-222). Springer, Cham.
- [29] Wu J, Wang Z. A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory. *Water*. 2022 Feb 17;14(4):610.
- [30] Than NH, Ly CD, Van Tat P. The performance of classification and forecasting Dong Nai River water quality for sustainable water resources management using neural network techniques. *Journal of Hydrology*. 2021 May 1;596:126099.
- [31] Yan J, Liu J, Yu Y, Xu H. Water quality prediction in the Luan River based on 1-DRCNN and bigru hybrid neural network model. *Water*. 2021 Apr 30;13(9):1273.
- [32] Sha J, Li X, Zhang M, Wang ZL. Comparison of forecasting models for real-time monitoring of water quality parameters based on hybrid deep learning neural networks. *Water*. 2021 May 31;13(11):1547.
- [33] J.Brownlee. Extreme Gradient Boosting (XGBoost) Ensemble in Python, *Machine Learning Mastery* (27/4/2021), [online], available: <https://machinelearningmastery.com/extreme-gradient-boosting-ensemble-in-python/> (accessed 22/4/2022).
- [34] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* 2016 Aug 13 (pp. 785-794).
- [35] J.Brownlee, How to Develop a Light Gradient Boosted Machine (LightGBM) Ensemble, *Machine Learning Mastery* (25/11/2020), [online], available at: <https://machinelearningmastery.com/light-gradient-boosted-machine-lightgbm-ensemble/> (accessed 22/4/2022).
- [36] Learn website, Decision Trees [online], available at: [https://scikit-learn.org/stable/modules/tree.html#:~:text=Decision%20Trees%20\(DTs\)%20are%20a,as%20a%20piecewise%20constant%20approximation.](https://scikit-learn.org/stable/modules/tree.html#:~:text=Decision%20Trees%20(DTs)%20are%20a,as%20a%20piecewise%20constant%20approximation.) (accessed 23/4/2022).
- [37] A.Gupta, ML | Extra Tree Classifier for Feature Selection, [online], available at: <https://www.geeksforgeeks.org/ml-extra-tree-classifier-for-feature-selection/> (accessed 23/4/2022).
- [38] Swarnimrai, Multi-Layer Perceptron Learning in Tensorflow, [online], available: <https://www.geeksforgeeks.org/multi-layer-perceptron-learning-in-tensorflow/?ref=gcse> (accessed 23/4/2022).
- [39] Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*. 2013 Dec 4;7:21. [online], available at: <https://www.frontiersin.org/articles/10.3389/fnbot.2013.00021/full> (accessed 10/5/2022).
- [40] Geeks for Geeks website, Support Vector Machine Algorithm, [online], available at: <https://www.geeksforgeeks.org/support-vector-machine-algorithm/?ref=gcse> (accessed 10/5/2022).
- [41] Kadam AK, Wagh VM, Muley AA, Umrikar BN, Sankhua RN. Prediction of water quality index using artificial neural network and multiple linear regression modeling approach in Shivganga River basin, India. *Modeling Earth Systems and Environment*. 2019 Sep;5(3):951-62.
- [42] Maulud D, Abdulazeez AM. A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*. 2020 Dec 31;1(4):140-7.

# Leaf-based Classification of Important Indigenous Tree Species by Different Feature Extraction Techniques and Selected Classification Algorithms

Eugene Val D. Mangaoang<sup>1</sup>, Jaime M. Samaniego<sup>2</sup>

Department of Computer Science & Technology, Visayas State University, Baybay City, Philippines<sup>1</sup>  
Institute of Computer Science-University of the Philippines-Los Baños College, Laguna, Philippines<sup>2</sup>

**Abstract**—The machine learning algorithms, namely, *k*-Nearest Neighbor (KNN), Support Vector Machine (SVM), Back-Propagation (BP) networks, and Convolutional Neural Networks (CNN) are four of the mostly used classifiers. Different sets of features are required as input in different application domains. In this paper, a set of significant leaf features and classification model was determined with a high accuracy in classifying important indigenous tree species. Leaf images were acquired using a scanner to control the image quality. The image dataset was then duplicated into two sets. The first set was labeled with their correct classes, preprocessed, and segmented in preparation for feature extraction. The leaf features extracted were leaf shape, leaf color, and leaf texture. Then, training and classification was done by KNN, SVM, and BP networks. On the other hand, the second set was unlabeled for training and classification by CNN. A CNN model was built and chosen with the best training and validation accuracy and the least training and validation loss rate. The study concluded that using all three leaf features for classification by BP networks resulted in a 93.48% accuracy with training done by supervised learning. However, the CNN achieved a high accuracy rate of 98.5% making it the best approach for classification of tree species using digital leaf images in the context of this study.

**Keywords**—Machine learning; feature extraction; convolutional neural network; leaf classification

## I. INTRODUCTION

Computer vision is divided into image acquisition, image preprocessing, feature extraction and description, and classification. In the case of plant species classification, an image of a part of a plant is acquired using a camera. The image is then preprocessed in preparation for the next step. Operations include eliminating noise, correcting geometric distortions or degraded image data, and segmentation. The aim is to emphasize the features of the image that are relevant for further processing while suppressing the undesired distortions. After image preprocessing, features are then extracted based on the descriptors established. These descriptors are a set of numbers that describe the part of the plant in the captured image. Finally, the plant species is recognized using all the features that have been extracted [1].

Feature extraction is a vital stage in image-based classification when it comes to the accuracy and precision of the classifier. This is because the underlying machine learning principles utilize the features that are supplied into the network.

Feature extraction techniques need to be selected thoroughly so that the image is well-perceived hence, providing the classifier enough information for a more accurate and precise classification [2].

After all the features of the subject in the image have been understood, classification is done by a mathematical classifier. There are numerous classification algorithms and each one requires a different set of features as input. Among the mostly used classifiers for image processing are *k*-Nearest Neighbor (KNN) algorithm, Support Vector Machine (SVM), Back-Propagation (BP) networks, and Convolutional Neural Networks (CNN) [2, 3, 4, 5, 6, 7, 8, 9]. In this study, the accuracy of classification of each of the algorithms were measured to determine which is best for leaf-based tree species classification.

Leaves are the suitable plant organs for computerized plant classification as these are numerous and acquired for most of the year rather than flowers and fruits since plants produce these in a limited period [10]. This is especially true for trees since it may take years to produce flowers and fruits. The bark of a tree is another organ aside from its leaves that can be used for tree recognition. Although these two organs are present throughout the year, tree recognition based on its bark can be very difficult and can add confusion [11]. In this study, leaf-based tree species classification was done using computer vision techniques.

As mentioned previously, the classification algorithms, KNN, SVM, BP networks, and CNN, require different sets of features extracted from an image for an accurate classification. These sets of features will always differ for different application domains. In the case of leaf-based tree species classification, there is a need to determine the significant features to be extracted so that there is a high accuracy in classifying tree leaf species.

The general objective of the study aimed to determine which combination of leaf features and classification algorithm provides an accurate and consistent classification of tree species through tree leaf images. Specifically, this study aimed to: (1) extract features from digital images of leaves of important indigenous tree species using image processing techniques, (2) implement classification algorithms for tree species classification, and (3) evaluate the accuracy of each of the classification algorithms.

Tree species classification has always been essential for understanding biodiversity conservation. Forests function well if there is a diversity of trees as they gather nutrients, which are then released when the trees die and decompose. These functions include growing wood to be used for furniture and timber among others, counteract climate change, and prevent soil erosion and regulate the water cycle. Forests that have diverse trees should provide better ecosystem goods and services to humans than those that have monocultured trees [12]. Scientists working to document and study forest flora and fauna, which is key to biodiversity conservation, are overwhelmed with the rich forest species biodiversity due to limited taxonomic expertise making it very difficult to abate the rapid degradation of forests [13]. On the other hand, inexperienced persons will even find it more difficult to classify the surrounding trees. Hence, the development of a program with enough tree taxonomic knowledge that enables recognition of tree species in a quick manner is significant to overcome this challenge. With the proposed tree recognition system, users will now have a tool that can classify trees. Users will be given good information and are likely able to make informed choices about which trees are suitable for biodiversity conservation.

This paper continues first by investigating which leaf features are commonly used in leaf classification in previous works. Then, KNN, SVM, BP networks, and CNN are also investigated in previous works. The performance of these classification algorithms against datasets in previous works were noted as well as the leaf features used for leaf classification. Next, the paper explains how the dataset was prepared for supervised and unsupervised learning and for classification. The results are then discussed, and the accuracy of the classifiers are indicated. The paper then concludes which classification algorithm worked best in the context of this study, and recommendations are also mentioned for future work.

## II. RELATED WORKS

### A. Leaf Features

There are many techniques that have been introduced to extract features from digital leaf images. Leaf shape is the most common extracted feature for classification. It is often used because of significant differences between leaf shapes of different plants. Important indigenous tree species in Leyte exhibit differences in their leaf shapes as well. Among the various works on leaf classification, the most common extracted representations of shape features are the following: aspect ratio, eccentricity, compactness / roundness, and rectangularity [14, 15, 16].

In some research, the leaf's texture is also considered in its classification [14, 17, 18, 19]. This is also an effective way of differentiating leaves as some have different textures compared to others, especially at a very small scale. The grey level co-occurrence matrix (GLCM) is commonly used to extract the leaf texture features. There are 14 textural features suggested to be computed [20]. These statistical features determine differences in the intensity of the pixels and the spatial relations between them.

Leaf color can also be used for plant species classification especially when there is a difference in leaf color among different species [17, 18]. Leaf color features are usually extracted along with another attribute, usually leaf shape and texture, when there is a subtle difference between leaf color among the species. Leaf color features are represented by the distribution of color in the image in each of the three color planes (red, green, and blue). If an image follows a certain distribution, then the color features can be used to identify the image.

### B. Mathematical Classifiers

This study assessed which of the classification algorithms, namely  $k$ -Nearest Neighbor, Support Vector Machine, Back-Propagation networks, and Convolutional Neural Networks would provide the most accurate results when classifying tree leaf images of important indigenous tree species. Related research works regarding the use of the algorithms are investigated. The commonly extracted features utilized by the classifiers are also taken into consideration.

1) *k*-Nearest Neighbor (KNN) algorithm: The author in [15] implemented the  $k$ -Nearest Neighbor (KNN) classifier for tree taxonomy. The leaf's shape features were extracted. It was concluded that the classifier best performed when eccentricity, a moment invariant and maximal indentation depth were the features used. The performance of the KNN classifier was low when all the features were used. This was because some of the features were redundant and hindered the success rate. Also, the success rate improved to almost 90% when the user inspected more nearest neighbors (specifically, when the  $k$ -value was 10). The author in [16] used three shape features: slimness (aspect ratio), roundness (compactness), and dispersion for classifying leaves using the KNN algorithm. The Flavia dataset was used which consists of 32 classes of 1907 leaf images. The classifier performed best when the  $k$ -value was 3, with an overall accuracy of 94.37% in classification. The author in [14] combined leaf shape features and leaf texture features for classifying leaves using the KNN algorithm. The shape features included aspect ratio, rectangularity, narrow factor, circularity, and solidity. Meanwhile, contrast, homogeneity, correlation, and energy were the texture features extracted using grey level co-occurrence matrix (GLCM). The classification had an accuracy rate of 94%. The author in [6] used leaf venations in identifying selected dicot plant species. The leaf venations were represented as a graph and relevant graph metrics were computed. Each of the graph metrics of the plant species served as input to different classifiers including the KNN algorithm. The study also used the Support Vector Machine (SVM) and Back-Propagation (BP) networks, specifically, Multilayer Perceptron (MLP), mathematical classifiers which are discussed in the next sections. The KNN classifier achieved an accuracy of 21.64% while the SVM and MLP classifiers reached 24.85% and 23.65%, respectively. According to the authors, this was due to the low cardinality of the dataset which was only ten leaves for each of the 50 plant species.



2) *Back-Propagation (BP) networks*: In [21], the accuracy rate of back-propagation (BP) neural networks and *k*-nearest neighbors (KNN) were compared. It was concluded that the BP networks produced a higher accuracy rate of 93.3% than the KNN, which had an accuracy rate of 85.9%, for a large dataset. For an otherwise small dataset, the KNN classifier outperforms the BP networks approach. The author in [10] also used BP networks in leaf-based classification of plants. The classifier produced 96% accuracy rate in classification. Both studies used the Flavia dataset and used shape features for classification.

3) *Support Vector Machine (SVM)*: The author in [18] utilized SVM for plant leaf recognition. The authors extracted the leaf texture and color features which were used by the classifier. With color features alone, the classifier produced low accurate results. This was because of the high similarity between the colors of the leaf images. But, when texture and color features were used, the accuracy rate of classification went up to 92%. The author in [17] also extracted leaf texture and color features for use of the SVM classifier. According to the authors, SVM performs well when compared to the KNN classifier. The system implemented in their work attained an average accuracy of 93.26%. The author in [19], on the other hand, only extracted texture features for classification using SVM. The classifier had an accuracy of 90.27%. The authors concluded that this may be improved if other leaf features could also be considered.

4) *Convolutional Neural Networks (CNN)*: In the study [7], CNN was employed for plant leaf recognition. The image dataset used was the Flavia dataset and the sizes were changed to 229x229 to fit the model. The structure of their proposed model had five convolution layers followed by their proposed inception module, then the pooling layer of size 8x8. Input images also included discolored leaves and damaged leaves. Despite the discoloration and damage of the leaves, the system has a recognition rate of above 94%. Another work in [8], uses a deep convolutional network model. It consists of 16 weight layers: 13 convolutional layers and 3 fully connected layers. The Flavia dataset was used, and the input images were resized into 224x224 pixels. Data augmentation was also performed which added the transformations of the initial image dataset as input. Because of this, using deep convolutional neural network, the system achieved an accuracy of 99.9%. The author in [9] also used data augmentation on the Flavia dataset. In their work, the images were resized to 256x256 pixels. There were three convolutional layers with an addition of a PReLU activation function after each convolution. The accuracy rate for the trained model is greater than 94.6% on 32 kinds of plants.

### III. METHODOLOGY

The tree species identified were limited to important indigenous tree species in Leyte due to the small number of tree classes which can all be classified through the tree's leaves. The 25 tree species to be classified are listed in Table I. The leaves of important indigenous trees were

collected from the forest reserve of the Visayas State University. The trees from which the leaves were collected have already matured spanning the ages between 15-20 years old. The framework shown in Fig. 1 depicts the different processes that were involved in the study which mainly applies image processing and the leaf classification techniques. Supervised learning is when the dataset was labeled. This was when KNN, SVM, and BP networks were used for training. Meanwhile in unsupervised learning, the dataset used was unlabeled and CNN was used for the training phase.

#### A. Image Acquisition and Preparation

With the help of an expert on taxonomic classification of important indigenous tree species, leaves were collected and scanned immediately using a scanner to control the image quality. The images were manually categorized according to its species and stored in their respective folders. A sequence of pre-processing techniques was performed that would make these images appropriate for extracting related information. The techniques included cropping, to emphasize the region of interest; scaling, to reduce the image size; and applying noise removal operation, to improve the image quality.

TABLE I. LIST OF TREE SPECIES CLASSIFIED

Family Name	Scientific Name	Local Name
Anacardiaceae	Dracontomelon dao	Dao
Calophyllaceae	Calophyllum blancoi	Bitanghol
Dipterocarpaceae	Shorea astylosa	Yakal
Dipterocarpaceae	Hopea plagata	Yakal Saplungan
Dipterocarpaceae	Shorea contorta	White Lauan
Dipterocarpaceae	Shorea almon	Almon
Dipterocarpaceae	Shorea squamata	Mayapis
Dipterocarpaceae	Dipterocarpus grandiflorus	Apitong
Dipterocarpaceae	Shorea falciferoides	Yakal Yamban
Dipterocarpaceae	Shorea guiso	Guijo
Dipterocarpaceae	Hopea philippinensis	Gisok-gisok
Dipterocarpaceae	Dipterocarpus validus	Hagakhak
Dipterocarpaceae	Parashorea malaanonan	Bagtikan
Dipterocarpaceae	Shorea polysperma	Tanguile
Dipterocarpaceae	Dipterocarpus kerrii	Malapanau
Dipterocarpaceae	Hopea malibato	Yakal Kaliot
Euphorbiaceae	Securinega flexuosa	Anislag
Fabaceae	Afzelia rhomboidea	Tindalo
Fabaceae	Pterocarpus indicus	Narra
Fagaceae	Lithocarpus llanosii	Ulayan
Lamiaceae	Vitex parviflora Juss.	Molave
Lecythidaceae	Petersianthus quadrialatus Merr.	Toog
Myrtaceae	Xanthostemon verdugonianus Naves	Mangkono
Sterculiaceae	Pterospermum acerifolium Willd.	Bayog
Tiliaceae	Diplodiscus paniculatus Turcz.	Balobo

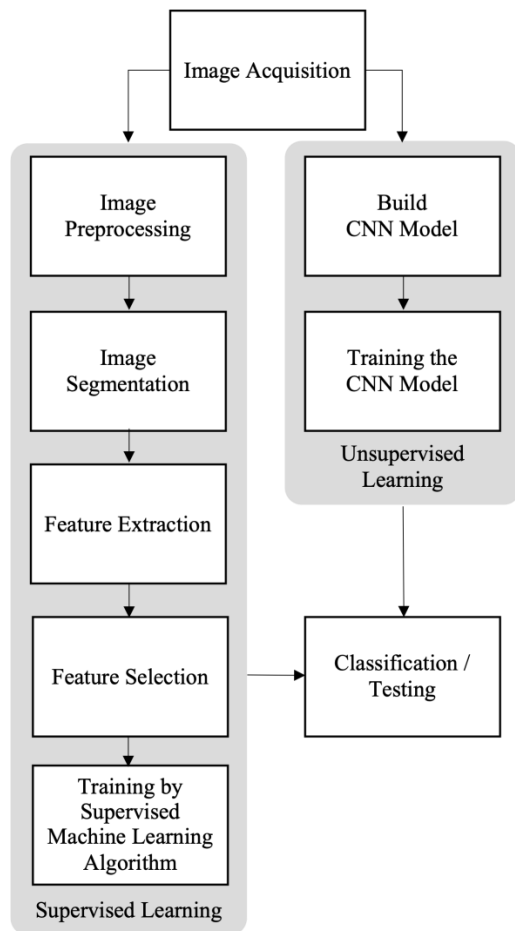


Fig. 1. Conceptual framework of the study.

### B. Image Segmentation

The leaf images were first converted into grayscale. Afterwards, Gaussian filter of size (25, 25) was applied to smooth the image. Next, adaptive image thresholding using Otsu's thresholding method was applied. Lastly, morphological closing was applied to close any holes present in the leaf. Fig. 2 shows how the image was transformed in preparation for feature extraction.

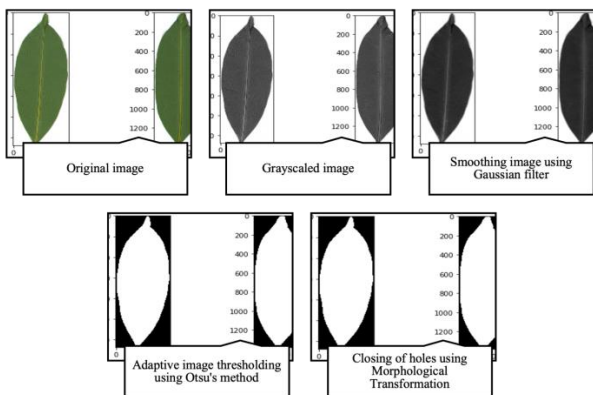


Fig. 2. Image segmentation techniques.

### C. Feature Extraction

Feature extraction was performed to get useful information that served as basis for leaf classification. Shape, color, and texture features were extracted from the leaf images. Boundary extraction was first done before calculating the shape features. Leaf boundary extraction was done using contours. The shape features that were calculated from the leaf images are area, roundness, aspect ratio, eccentricity, and rectangularity. For leaf color feature extraction, the original colored image was converted from RGB to HSV to get the hue, saturation, and value. The mean, standard deviation, skewness, and kurtosis of the HSV values were calculated. The texture features were extracted using the grey level co-occurrence matrix (GLCM). The 14 Haralick textural features (angular second moment, contrast, correlation, etc.) were computed. The original colored image was converted from RGB to grayscale before computing for the 14 features.

### D. Feature Selection

All the extracted features, especially the texture features, may be correlated to each other. This means that there may be redundant features that do not contribute information for classification. So, feature selection was employed to find the most suitable features to improve the accuracy of classification. Feature selection was done using Pearson's correlation through the Waikato Environment for Knowledge Analysis (Weka) tool.

### E. Training and Classification Phase

The training dataset was composed of 80% of the total number of acquired images. To determine which set of leaf features was best for classification, shape, texture, and color features were first used individually for training and classification then, a combination of two features, and finally, all three leaf features were used as basis for classification. The KNN, BP, and SVM classification algorithms were used for training and classification of the tree leaf species through the Weka tool.

For training and classification through CNN, TensorFlow was used to build the classification model. 80% of the image dataset was also used for training and 10% of the image dataset was used for validation. The mini-batch gradient descent learning algorithm was applied on the dataset with 32 batches processed at each time. This means that 789 batches per epoch were processed to go over the 25,271 images. The learning process went for 250 epochs and the validation dataset was used as reference to determine the performance of the model for each epoch.

When compiling the model, the standard cross-entropy loss was used to calculate the error rate of prediction from the original value. Categorical class classification was used to predict from 25 tree species/classes. The adam optimizer was used to adjust how the model learns during the training process. Finally, the accuracy was used to determine how the model can correctly predict the tree species using the validation dataset during training phase.

F. Testing Phase

The remaining 20% of the total number of images was used as test dataset. The same feature extraction processes were performed on the dataset and then classification was carried out by the classifiers. For testing the best CNN model, 10% of the image dataset was used. The accuracy of the classifiers was computed using (9):

$$\text{Accuracy} = \frac{\text{number of correctly classified images}}{\text{total number of testing images}} \times 100 \quad (1)$$

In addition, the following evaluation metrics for classification models were also computed:

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \times 100 \quad (2)$$

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \times 100 \quad (3)$$

$$\text{Specificity} = \frac{\text{True Negatives (TN)}}{\text{True Negatives (TN)} + \text{False Positives (FP)}} \times 100 \quad (4)$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \quad (5)$$

G. Building the CNN Model

The CNN architecture used in this study, shown in Fig. 3, was inspired by a similar architecture albeit with modifications to the parameters and configurations used in [22]. The dataset first passes through convolution layers followed by activation layers, and then pooling layers. This sequence of layers is repeated twice to add more hidden layers. Lastly, the fully connected network acts as the classifier for the model.

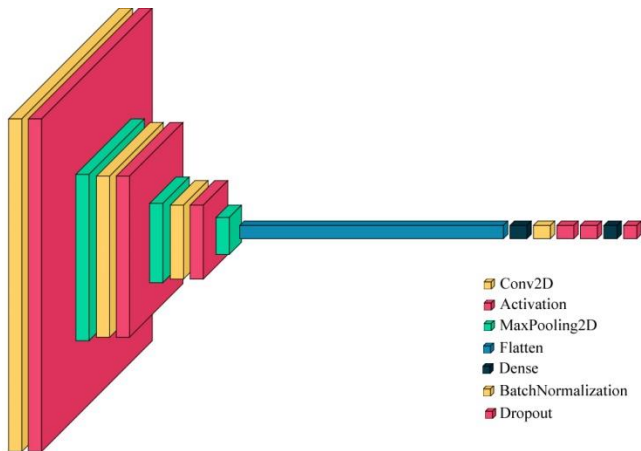


Fig. 3. Convolutional neural network architecture of the study.

The convolutional layers extract features from the input images. The number of filters follows the trend of 32-64-128 filters. Rectified Linear Unit (ReLU) activation was used for the activation layers. Pooling layers have a filter size of 2x2 with a stride of two which shrinks the dimensions of the data by half. After passing through the convolution-activation-pooling sequence twice, the final output was flattened into a vector and fed into a fully connected dense network. The first dense layer has 256 nodes followed by a batch normalization layer which standardizes the input before being activated by a ReLU function. Dropout layer follows next to prevent

overfitting. The last dense layer has 25 nodes activated by softmax activation layer which allows the model to predict from 25 tree species with the highest probability. Table II depicts the layers and their corresponding output shape as well as the number of parameters.

TABLE II. SUMMARY OF THE MODEL

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 126, 126, 32)	896
activation (Activation)	(None, 126, 126, 32)	0
max_pooling2d (MaxPooling2D)	(None, 63, 63, 32)	0
conv2d_1 (Conv2D)	(None, 61, 61, 64)	18496
activation_1 (Activation)	(None, 61, 61, 64)	0
max_pooling2d_1 (MaxPooling2D)	(None, 30, 30, 64)	0
conv2d_2 (Conv2D)	(None, 28, 28, 128)	73856
activation_2 (Activation)	(None, 28, 28, 128)	0
max_pooling2d_2 (MaxPooling2D)	(None, 14, 14, 128)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 256)	6422784
batch_Normalization (BatchNormalization)	(None, 256)	1024
activation_3 (Activation)	(None, 256)	0
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 25)	6425
activation_4 (Activation)	(None, 25)	0
Total params:	6,523,481	
Trainable params:	6,522,969	
Non-trainable params:	512	

IV. RESULTS AND DISCUSSIONS

The total number of leaf images scanned is 31,508. This includes both the top and underneath sides of the leaf. The distribution of the images per tree species is shown in Table III.

TABLE III. NUMBER OF IMAGES PER TREE SPECIES

Class	No. of images	Class	No. of images
almon	1258	mayapis	1184
anislag	1264	molave	1212
apitong	1112	narra	1328
bagtikan	1308	tanguile	1284
balobo	1300	tindalo	1404
bayog	1282	toog	1208
bitanghol	1240	ulayan	1300
dao	1526	white lauan	1158
gisok-gisok	1236	yakal	1164
guijo	1386	yakal kaliot	1224
hagakhak	1252	yakal saplungan	1198
malapanau	1234	yakal yamban	1216
mangkono	1230		

**A. Supervised Learning Results**

The shape, color, and texture features were computed through Python code and saved in a .csv file. These were then used for training and classification by KNN, SVM, and BP networks using Weka. Fig. 4 shows the accuracy of the classifiers with the corresponding leaf features as input.

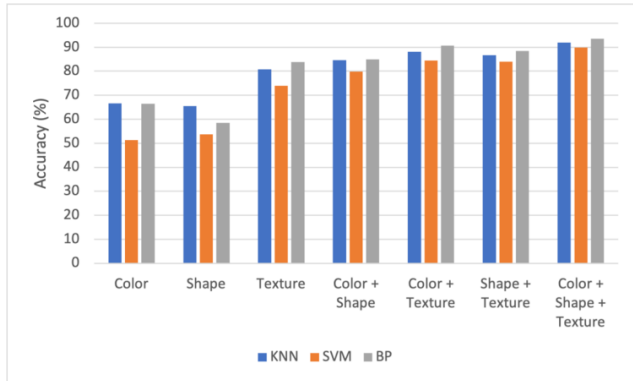


Fig. 4. Accuracy of classifiers through supervised learning.

The individual sets of leaf features produced relatively low values of accuracy, the lowest being only 51.3% using the leaf color features by the SVM classifier. As the number of leaf features used was added, the accuracy improved. Among the two sets of leaf features used, color and texture features produced the best accuracy by the three classifiers with BP classifier having an accuracy of 90.65%. With the color, shape, and texture features used, the KNN classifier displayed 91.873% accuracy, and the SVM classifier displayed 89.873% accuracy. Meanwhile, the BP classifier showed the best accuracy among the three with 93.4762%. Table IV shows the time taken to build the models with the corresponding combination of leaf features extracted.

TABLE IV. TIME (IN SECONDS) TO BUILD THE MODEL THROUGH SUPERVISED LEARNING

Features Extracted	KNN	SVM	BP
Color	0.02	4.98	147.67
Shape	0.01	2.39	115.01
Texture	0.01	6.36	265.35
Color + Shape	0.03	4.46	188.51
Color + Texture	0	6.8	403.79
Shape + Texture	0	5.9	306.68
All three features	0.02	6.63	463.54

Using the model built through back-propagation network, a confusion matrix was derived. Table V shows a high performance of the BP model in classifying tree species through leaf images with up to 99% precision.

**B. CNN Model Results**

The training and validation performance of the CNN model is shown in Table VI. It took an accumulated time of about 50 hours to finish the training and validation phase due to the numerous digital leaf images. For choosing the best CNN model, it is generally ideal to choose a model with the best training and validation accuracy as well as the least

training and validation loss rate. The best training accuracy rate is 98.62% at epoch 240 while the least loss value is 0.0431 at epoch 245. However, the best model can be observed at epoch 249, with a significant validation accuracy rate of 98.79% and a loss value of 0.0554.

TABLE V. EVALUATION METRICS OF THE BP MODEL

Class	Precision	Recall	Specificity	F1 Score
almon	97.08%	92.46%	99.88%	94.72%
anislag	98.37%	48.02%	99.97%	64.53%
apitong	99.10%	98.21%	99.97%	98.65%
bagtikan	82.76%	92.31%	99.17%	87.27%
balobo	99.62%	99.62%	99.98%	99.62%
bayog	96.12%	96.88%	99.83%	96.50%
bitanghol	95.74%	99.60%	99.82%	97.63%
dao	93.22%	90.46%	99.67%	91.82%
gisok-gisok	84.59%	95.16%	99.29%	89.56%
guijo	96.30%	94.20%	99.83%	95.24%
hagakhak	91.54%	98.81%	99.62%	95.04%
malapanau	97.48%	93.55%	99.90%	95.47%
mangkono	97.00%	91.13%	99.88%	93.97%
mayapis	90.46%	92.37%	99.62%	91.40%
molave	89.59%	98.77%	99.54%	93.96%
narra	91.04%	92.42%	99.60%	91.73%
tanguile	93.98%	97.66%	99.74%	95.79%
tindalo	93.20%	97.86%	99.67%	95.47%
toog	84.36%	96.67%	99.29%	90.10%
ulayan	91.82%	95.00%	99.64%	93.38%
white lauan	89.71%	94.37%	99.59%	91.98%
yakal	100.00%	97.42%	100.00%	98.70%
yakal kaliot	97.56%	98.36%	99.90%	97.96%
yakal saplungan	96.76%	87.08%	99.88%	91.67%
yakal yamban	97.98%	99.59%	99.92%	98.78%

TABLE VI. TRAINING AND VALIDATION LOSS AND ACCURACY OF THE CNN MODEL

Epoch No.	Training		Validation	
	Loss	Accuracy	Loss	Accuracy
1	2.436	0.2602	1.8804	0.3785
25	0.2588	0.916	0.6342	0.8428
50	0.15	0.9507	0.6239	0.8597
75	0.1079	0.9663	0.3062	0.9228
100	0.0959	0.9701	0.1508	0.9563
125	0.0729	0.9765	0.2956	0.9391
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
240	0.0446	0.9862	0.1643	0.9601
245	0.0431	0.9857	0.1349	0.9688
249	0.0503	0.9843	0.0554	0.9879
250	0.0497	0.985	0.1343	0.9707

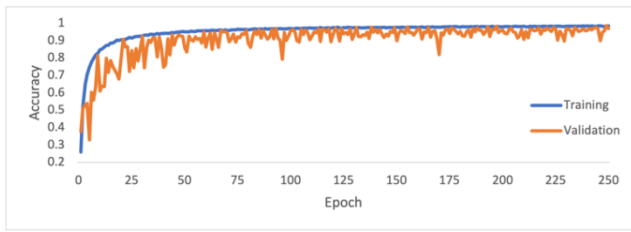


Fig. 5. Accuracy plot of the models by epoch

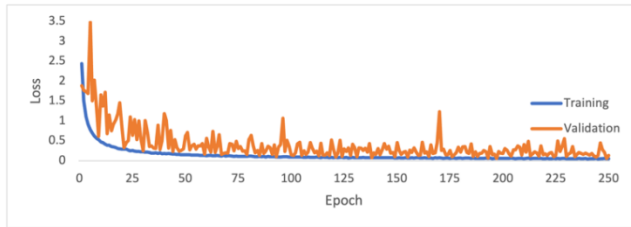


Fig. 6. Loss plot of the models by epoch

Fig. 5 shows the CNN model accuracy by epoch while Fig. 6 shows the CNN model loss values by epoch. It is demonstrated in Fig. 5 that the model has learned significantly at epochs 1-25 and the accuracy during training and validation continues to improve. It also shows the model’s improving performance in the validation phase is not significantly different from the training phase.

TABLE VII. EVALUATION METRICS OF THE CNN MODEL

Class	Precision	Recall	Specificity	F1 Score
almon	99.21%	100.00%	99.97%	99.60%
anislag	94.74%	100.00%	99.77%	97.30%
apitong	94.07%	99.11%	99.77%	96.52%
bagtikan	96.30%	100.00%	99.83%	98.11%
balobo	99.23%	99.23%	99.97%	99.23%
bayog	99.21%	97.66%	99.97%	98.43%
bitanghol	100.00%	100.00%	100.00%	100.00%
dao	100.00%	94.08%	100.00%	96.95%
gisok-gisok	98.35%	95.97%	99.93%	97.14%
guijo	100.00%	97.10%	100.00%	98.53%
hagakhak	98.43%	99.21%	99.93%	98.81%
malapanau	98.40%	99.19%	99.93%	98.80%
mangkono	100.00%	97.58%	100.00%	98.78%
mayapis	97.52%	100.00%	99.90%	98.74%
molave	100.00%	98.36%	100.00%	99.17%
narra	98.50%	99.24%	99.93%	98.87%
tanguile	100.00%	100.00%	100.00%	100.00%
tindalo	98.58%	99.29%	99.93%	98.93%
toog	96.75%	99.17%	99.87%	97.94%
ulayan	99.23%	99.23%	99.97%	99.23%
white lauan	98.31%	100.00%	99.93%	99.15%
yakal	100.00%	98.28%	100.00%	99.13%
yakal kaliot	100.00%	99.18%	100.00%	99.59%
yakal saplungan	100.00%	92.50%	100.00%	96.10%
yakal yamban	96.83%	100.00%	99.87%	98.39%

The trained CNN model at epoch 249 was used for classifying the testing image dataset. Out of 3,151 images, the model was able to correctly classify 3,104 leaves according to their tree species. Therefore, the model has an accuracy rate of 98.5%.

A confusion matrix was derived from classifying the 25 tree species using the CNN model. Evaluation metrics were derived from the confusion matrix as shown in Table VII. This reflects the model’s high performance in classifying the 25 tree species especially the bitanghol and tanguile tree species with a recognition rate of 100%.

## V. CONCLUSIONS AND RECOMMENDATIONS

The study was able to extract the leaf color, shape, and texture features from digital images of leaves of important indigenous tree species using image processing techniques. A combination of these features was used for classification by three machine learning algorithms: *k*-Nearest Neighbor algorithm, Backpropagation networks, and Support Vector Machine. Among the possible leaf features combinations, it shows that using all the three features provides higher accuracy of classification compared to using just one or a combination of two leaf features. BP networks also provides the highest accuracy with 93.48% out of the three supervised machine learning algorithms for this study. However, the model built using Convolutional Neural Network has an accuracy rate of 98.5% making it the best approach for classification of tree species using digital leaf images in the context of this study.

It is recommended to include more tree species for classification as well as adding more shape features, and even more leaf features like leaf venation, in the case of supervised learning. This may reduce the likelihood of Type I and Type II classification errors. Modifying the configurations in building the CNN model is also recommended to further increase the accuracy rate. The models could also be trained in order to recognize unknown classes or images that are outside of the training dataset. Additionally, using other devices to acquire more detailed leaf images, such as digital SLR cameras or hyperspectral imaging devices, could also be used to extract more features from leaves for a more accurate classification.

## ACKNOWLEDGMENT

The authors would like to thank Dr. Eduardo O. Mangaoang of the Visayas State University, for his expert knowledge on taxonomic classification of important indigenous tree species and for facilitating the leaf collection in the forest reserve of the Visayas State University. Heartfelt gratitude is also given to the authors’ friends and colleagues whose valuable insights contributed to this study’s success. This study was also supported by the DOST-ASTHRDP scholarship grant.

## REFERENCES

- [1] J. Wäldchen and P. Mäder, “Plant species identification using computer vision techniques: a systematic literature review,” *Archives of Computational Methods in Engineering*, vol. 25, no. 2, pp. 507-543, 2018.

- [2] M. A. F. Azlah, L. S. Chua, F. R. Rahmad, F. I. Abdullah and S. R. Wan Alwi, "Review on techniques for plant leaf classification and recognition," *Computers*, vol. 8, no. 4, p. 77, 2019.
- [3] K. Manjula, K. Vijayarekha and P. Vimaladevi, "Review on classification algorithms in image processing," *International Journal of Innovative Trends in Engineering & Research*, vol. 2, no. 11, pp. 1-6, 2017.
- [4] S. Aggarwal and M. Bhatia, "Anatomy of leaf classification techniques," in *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing*, Faridabad, India, 2019.
- [5] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, 2012.
- [6] A. K. D. Balangcod and J. P. Pabico, "Automatic Identification of Selected Dicot Plant Species Using Graph Properties from Leaf Venations," in *Proceedings of Philippine Computing Science Congress*, City of Baguio, Philippines, 2020.
- [7] W.-S. Jeon and S.-Y. Rhee, "Plant leaf recognition using a convolution neural network," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 17, no. 1, pp. 26-34, 2017.
- [8] T. L. I. Sugata and C. K. Yang, "Leaf App: Leaf recognition with deep convolutional neural networks," *IOP Conference Series: Materials Science and Engineering*, no. 273, 2017.
- [9] C. Zhang, P. Zhou, C. Li and L. Liu, "A convolutional neural network for leaves recognition using data augmentation," in *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, Liverpool, UK, 2015.
- [10] A. Aakif and M. F. Khan, "Automatic classification of plants based on their leaves," *Biosystems Engineering*, vol. 139, pp. 66-75, 2015.
- [11] S. Bertrand, R. B. Ameer, G. Cerutti, D. Coquin, L. Valet and L. Tougne, "Bark and leaf fusion systems to improve automatic tree species recognition," *Ecological Informatics*, vol. 46, pp. 57-73, 2018.
- [12] S. Ratcliffe, C. Wirth, T. Jucker, F. van der Plas, M. Scherer-Lorenzen, K. Verheyen, E. Allan, R. Benavides, H. Bruelheide, B. Ohse, A. Paquette, E. Ampoorter, C. C. Bastias, J. Bauhus, D. Bonal, O. Bouriaud, F. Bussotti, M. Carnol, B. Castagneyrol, E. Češko, S. M. Dawud, H. D. Wandeler, T. Domisch, L. Finér, M. Fischer, M. Fotelli, A. Gessler, A. Granier, C. Grossiord, V. Guyot, J. Haase, S. Hättenschwiler, H. Jactel, B. Jaroszewicz, F.-X. Joly, S. Kambach, S. Kolb, J. Koricheva, M. Liebersgesell, H. Milligan, S. Müller, B. Muys, D. Nguyen, C. Nock, M. Pollastrini, O. Purschke, K. Radoglou, K. Raulund-Rasmussen, F. Roger, P. Ruiz-Benito, R. Seidl, F. Selvi, I. Seiferling, J. Stenlid, F. Valladares, L. Vesterdal and L. Baeten, "Biodiversity and ecosystem functioning relations in European forests depend on environmental context," *Ecology Letters*, no. 20, pp. 1414-1426, 2017.
- [13] A. H. Gentry, "Tropical forest biodiversity: distributional patterns and their conservational significance," *Oikos*, vol. 63, no. 1, pp. 19-28, 1992.
- [14] L. Hamrouni, O. Aiadi, B. Khaldi and M. L. Kherfi, "Plant species identification using computer vision techniques," *Revue des bioressources*, vol. 7, no. 1, pp. 113-120, 2017.
- [15] E. J. Pauwels, P. M. de Zeeuw and E. B. Rangelova, "Computer-assisted tree taxonomy by automated image recognition," *Engineering Applications of Artificial Intelligence*, vol. 22, pp. 26-31, 2009.
- [16] P. S. V. S. R. Kumar, K. N. V. Rao, A. S. N. Raju and D. J. N. Kumar, "Leaf classification based on shape and edge feature with k-NN classifier," in *2016 2nd International Conference on Contemporary Computing and Informatics*, Noida, India, 2016.
- [17] S. Kaur and P. Kaur, "Plant species identification based on plant leaf using computer vision and machine learning techniques," *Journal of Multimedia Information System*, vol. 6, no. 2, pp. 49-60, 2019.
- [18] Q. K. Man, C. H. Zheng, X. F. Wang and F. Y. Lin, "Recognition of plant leaves using support vector machine," in *Advanced Intelligent Computing Theories and Applications with Aspects of Contemporary Intelligent Computing Techniques*, Springer, Berlin, Heidelberg, pp. 192-199, 2008.
- [19] D. Puri, A. Kumar, J. Virmani and Kriti, "Classification of leaves of medicinal plants using Laws' texture features," *International Journal of Information Technology*, 2019.
- [20] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, Vols. SMC-3, no. 6, pp. 610-621, 1973.
- [21] V. Satti, A. Satya and S. Sharma, "An automatic leaf recognition system for plant identification using machine vision technology," *International Journal of Engineering Science and Technology*, vol. 5, no. 4, pp. 874-879, 2013.
- [22] J. F. V. Oraño, E. A. Maravillas and C. J. G. Aliac, "Jackfruit Fruit Damage Classification using Convolutional Neural Network," *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, pp. 1-6, 2019.

# A Novel Deep-learning based Approach for Automatic Diacritization of Arabic Poems using Sequence-to-Sequence Model

Mohamed S. Mahmoud<sup>1</sup>, Nermin Negied<sup>2</sup>

Technical University of Munich, School of Informatics, Germany<sup>1</sup>

School of Communication and Information Engineering, Zewail City, Giza, Egypt<sup>2</sup>

School of Engineering and Applied Science, Nile University, Giza, Egypt<sup>2</sup>

**Abstract**—Over the last 10 years, Arabic language have attracted researchers in the area of Natural Language Processing (NLP). A lot of research papers suddenly emerged in which the main work was the processing of Arabic language and its dialects too. Arabic language processing has been given a special name ANLP (Arabic Natural Language Processing). A lot of ANLP work can be found in literature including almost all NLP applications. Many researchers have been attracted also to Arabic linguistic knowledge. The work expands from Basic Language Analysis to Semantic Level Analysis. But Arabic text semantic analysis cannot be held without considering diacritization, which can greatly affect the meaning. Many Arabic texts are written without diacritization, and Diacritizing them manually is a very tiresome process that may need an expert. Automatic diacritization systems became a demand as an initial step for processing Arabic text for any Arabic Language Processing application as Arabic diacritization is very important to get a readable and understandable Arabic text. For this reason, many researchers recently worked on building systems and tools that automatically diacritize un-diacritized Arabic texts. This work presents a novel deep learning-based sequence-to-sequence model to diacritize un-diacritized Arabic poems. The proposed model was tested and achieved high diacritization accuracy rate.

**Keywords**—Text diacritization; deep learning; sequence-to-sequence; regex; tokenization; ANLP

## I. INTRODUCTION

Arabic language is spoken by more than 360 million people, and its speakers are distributed in the Middle East, and many other neighboring regions. It is noteworthy to know that it is one of the six official languages of the United Nations. Some ANLP work in literature includes sentiment analysis [1], machine translation [2], dialect identification [3], Named Entity Recognition NER [4], question answering [5], summarization [6], etc. What makes Arabic special and challenging at the same time is the diacritization. Diacritics play an important role for understanding the meaning of Arabic statements. Undiacritized Arabic sentence may be ambiguous and difficult to understand. Arabic text is naturally written in some percentage of diacritics to avoid misleading meaning. The percentage of these diacritics depends on the context or the domain. For example, religious text like the

Holy Quran is fully diacritized to minimize chances of understanding it incorrectly. Hadeeth books are also diacritized with high percentage. So are kids' educational books. Classical literature and poetry tend to be partially diacritized as well. However, newspapers and other genre are rarely diacritized. Naturally Arabic text consists of two classes of symbols: letters and diacritics. Letters comprise long vowels such as A, y, w as well as consonants. Diacritics on the other hand comprise short vowels, gemination markers, nunation markers, as well as other markers (such as hamza, the glottal stop which appears in conjunction with a small number of letters, dots on letters, elongation, and emphatic markers) which in all, if present, render an exact precise reading of a word. The remaining of this paper is organized as follows: A literature survey of the recent work in this area can be found in the following section. Section III describes the approach proposed by this paper. Experimental work is demonstrated in Section IV. Section V presents the Results and Discussion. Conclusion and suggestions for future work can be found in Section VI.

## II. LITERATURE REVIEW

A lot of Arabic Natural Language Processing ANLP work has been done in literature. For instance, Farghaly and Shaalan started their investigations in 2009. Their research showed the significance of the Arabic language and its properties. Their work also discussed the challenging characteristics of the language like agglutination and morphological non-concatenation [7]. Habash in 2010 [8] has wrote a comprehensive survey about Arabic language characteristics and features. The survey wasn't limited only to Modern Standard Arabic (MSA) but also, it's dialects. Habash's survey extended to talk about different orthographic, morphological and syntactic aspects of this language in details. Five years later another ANLP comprehensive survey of Shoufan and Alameri [9] has been published. The authors proposed a general categorization of dialectical Arabic language processing into four main classes: (1) Basic Language Analyses (BLA). (2) Building Resources (BR), (3) Language Identification (LI) and (4) Semantic-Level Analysis (SemA). Their survey can be considered as a reference for knowing relevant contributions that address a specific ANLP aspect for Arabic dialects.

Moving to automatic Arabic diacritization attempts found in literature, Diab and Habash [10] studied the impact of Arabic diacritization on statistical machine translation (SMT). Their research confirmed that the SMT performance is positively affected by the presence of diacritization to the extent that would make it robust to slightly high OOV rates. They reported that diacritization step is needed before translation attempts to achieve high SMT performance. In 2008, Shaalan et al [11] used SVM to diacritize undiacritized Arabic text and they achieved accuracy rate of 95.3% and 82% F-measure, but their technique only adds diacritics to the last letter of the word, while the diacritization of internal letters can affect the meaning, for example عمان means (Amman – Capital of Jourdan) and عُمان (Oman - Arabian Country). In 2016, Bouamor et al [12] limited their study to three types of diacritical marks: short vowels, nunation, and shadda (gemination). They were mainly interested in reducing the time consumed in Diacritizing undiacritized Arabic text, so they conducted a pilot study for a minimum diacritization scheme. Their proposed scheme mainly encoded the most relevant differentiating diacritics to reduce confusability among words that are homographs but not homophones. They finally confirmed that it is difficult to build such a scheme because of subjectivity and they promised to dig deeper in future work. In 2017, Darwish et al [13] built an Arabic diacritizer using a Viterbi decoder to diacritize every word including stems and all morphological patterns. Their diacritizer also used transliteration mining combined with sequence labeling to diacritize named entities that have English transliterations. They also used SVM but combined with some filters, and they achieved good results for case ending diacritization. In the same year, Fashwan and Alansari [14] used morphological and syntactic processing to discretize Arabic texts, in which Morphology-dependent that selects the best internal diacritized form of the same spelling, whereas Syntax-dependent that detects the best syntactic case of the word within a given sentence using the parsing tree of that sentence. In their discussion section they stated that they need to improve their system through adding more morphological and syntactic Arabic rules. In 2018, Alosaimy et al [15] have created a tool for accurate diacritization of highly cited texts by automatically “borrowing” diacritization from other citations using n-grams matching, and they used the well-known sunnah book ‘Riyad Al Salheen’ to test their tool. They reported a high accuracy in their conclusion, but their tool only works with highly cited texts. Darwish et al [16] in the same year, limited their study to Moroccan and Tunisian Arabic, using deep neural networks DNN, and they reported a small error rate for both languages. In 2019, Fadel et al [17], used Feed-Forward Neural Network (FFNN) and Recurrent Neural Network (RNN), to diacritize Arabic texts. Their models were trained using 50 epochs and tested only on the available benchmark datasets, and they reported a good diacritization results. M. Madhfar, A. M. Qamar in 2021 [18], used simple baseline deep learning model, encoder-decoder model, and encoder model that is used in text-to-speech system. The authors mainly studied the impact of presence and

absence of diacritics on a text-to-speech, and they concluded that diacritics play a major role in the successfulness of almost all ANLP applications. They also stated that there is a big difference between Modern Standard Arabic (MSA) and Classical Arabic (CA) language, the fact that makes the problem more challenging and data demanding. Abuali and Kurdy in 2022 [19], studied the relationship between the full diacritization of the Arabic text and the quality of the speech synthesized in screen readers, and they came up with a conclusion that it is mandatory to diacritize Arabic text before converting it to audio. Thompson and Alshehri [20] in 2022, stated that training a diacritizer to diacritize and translate Arabic text is much better than training the model to only diacritize the text as it resolves a lot of ambiguity issues. The authors used the standard data splits of Diab et al. [21], and they achieved a low error rate of 4.8% on the Penn Arabic Tree-bank datasets.

### III. DATASET

In this work, we collected a variety of textual sentences along with their diacritized version. The sentences varied in lengths and contexts (talking about different topics) in Arabic Language. Fig. 1 depicts the histogram of the different sentence lengths in our data (number of characters is used to measure sentence length). As you can see from the figure, nearly all the data are concentrated at lengths less than half of the maximum length in both the un-diacritized (Yellow) and diacritized sentences (Blue) which is a good reason to use the half of the maximum length as a padding length in the data preprocessing stage. More about the padding and data preprocessing will be explained in the next section.

The collected data contains 2000 sentences being represented as an ordered set of characters in the input and their equivalent 2000 sentences characters and their diacritization in the output Table I. This research, unlike related work in literature covers all types of diacritization of Arabic language (Harakat) such as: Fataha, Kasara, Damma, Skoon, Fathatan, Kasratan, Dammatan and Shaddah.

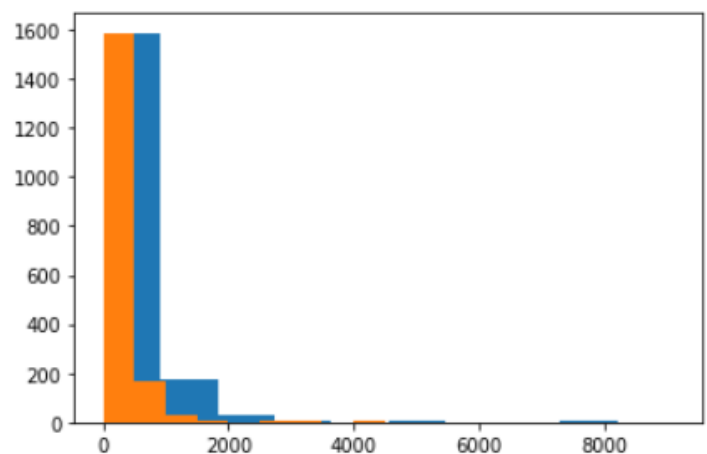


Fig. 1. Histogram of the lengths of the sentences diacritized (Blue) and un-diacritized (Yellow).



TABLE I. EXAMPLE OF INPUT / OUTPUT TEXT SEQUENCE

<p><b>Input Sequence</b></p>	<p>كذلك كان من أهداف حركة الاستشراق خدمة مخططات اليهود في هدم الإسلام والتمكين لهم في فلسطين عن طريق تشويه التاريخ العام، وتأكيد حق اليهود في فلسطين، وذلك ما يشير إليه المرحوم الدكتور نجيب البهي بقوله: إن هؤلاء أقبلوا على الاستشراق لأسباب دينية، وهي محاولة إضعاف الإسلام والتشكيك في قيمته وإثبات فضل اليهود على الإسلام بادعاء أن اليهودية هي مصدر الإسلام الأول، ولأسباب سياسية تتمثل بخدمة الصهيونية فكرة أول، ثم دولة ثاني. ولا يغيب عن بالنا أن المستشرقين هم الذين طرحوا على إنجلترا فكرة إنشاء وطن قومي لليهود في فلسطين في مؤتمر لندن المنعقد في سنة سبع وتسعمئة وألف</p>
<p><b>Output Sequence</b></p>	<p>كذلك كان من أهداف حركة الاستشراق خدمة مخططات اليهود في هدم الإسلام والتمكين لهم في فلسطين عن طريق تشويه التاريخ العام، وتأكيد حق اليهود في فلسطين، وذلك ما يشير إليه المرحوم الدكتور نجيب البهي بقوله: إن هؤلاء أقبلوا على الاستشراق لأسباب دينية، وهي محاولة إضعاف الإسلام والتشكيك في قيمته وإثبات فضل اليهود على الإسلام بادعاء أن اليهودية هي مصدر الإسلام الأول، ولأسباب سياسية تتمثل بخدمة الصهيونية فكرة أول، ثم دولة ثاني. ولا يغيب عن بالنا أن المستشرقين هم الذين طرحوا على إنجلترا فكرة إنشاء وطن قومي لليهود في فلسطين في مؤتمر لندن المنعقد في سنة سبع وتسعمئة وألف</p>

IV. THE PROPOSED APPROACH

Using our collected diacritized dataset, the proposed methodology exploits techniques in Deep Learning and Text Analysis to diacritize un-diacritized Arabic text. The objective is to take an input un-diacritized text and output its equivalent diacritized text. Unlike most of work in literature that used regular text analysis or statistical text classification to partially diacritize un-diacritized Arabic text, we used sequence-to-sequence model to diacritize text through mapping the input sequence of un-diacritized characters into diacritized ones. In our work, we clean the input and output texts and generate a parallel map between the two texts as a first stage to be able to train the deep learning model as a second stage. A further explanation of both stages can be found in the following subsections:

A. Data Preprocessing

In the data preprocessing stage, several text processing techniques were utilized. At first, data cleaning was used to remove irregular characters from the data such as “\ufeff” and “\u200f” which appeared so often because the data was stored in a bytes-like fashion. Then, regular expressions were used to remove any irregular character to Arabic language from the data including but not limited to English characters, question marks, dashes, dollar signs, asterisks, etc. Fig. 2 shows the regex we used to remove all the strange characters from the data. After that we tokenized the input and output text into their letters and diacritization (character wise tokenization). The character tokenization was done to make the model predicting the diacritic per character not per word. Following tokenization, we padded the sequences to a maximum length which we chose to be half of the maximum length and because of limitations on computational power. After the padding, we extracted all the unique tokens from the

input text and output diacritized text which was 36 (Arabic characters count) and 44 characters (Characters + diacritics) respectively. The strength of the regex we used was the main reason to be able to extract the unique characters of the data and to make it ready for use. We then used these unique characters to do one hot encoding of the data to be able to use them for deep learning training. Fig. 3 depicts the steps of the text preprocessing stage.

```
Regex [?!$%&'()*+,-./:;<=>@|~|^|_|`|{|}|\xa0|/|_|-|-|-]
```

Fig. 2. Regex symbols used in this work to clean data.

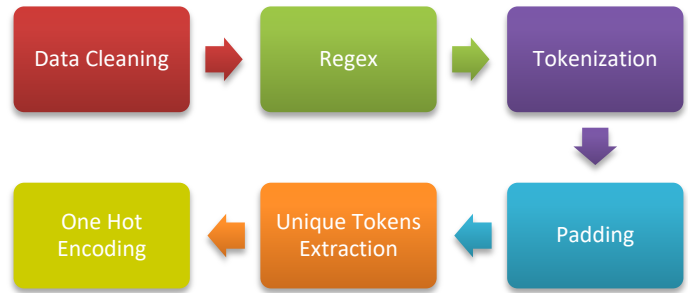


Fig. 3. The basic steps of text preprocessing stage.

B. Sequence-to-Sequence LSTM Model

After the text preprocessing stage, a sequence-to-sequence language model was built. The input to the model was the padded sequence one hot encoding of characters obtained from the text preprocessing stage. The sizes of the input data array and output data array were (4000, 4994, 36) and (4000, 9124, 44) respectively. The 4000 represents the total number of sentences we had that is why it is fixed in both input and output arrays. 4994 and 9124 indicate the padding length of the inputs and outputs respectively (half of the maximum length of both). Lastly, the 36 and 44 represent the number of unique tokens in the input and output respectively. The sequence-to-sequence model can generally be seen as encoder-decoder model. The encoder model has an input layer and an LSTM layer of 200 nodes. The decoder model has also input layer, LSTM layer of 200 nodes and output dense layer of the 44 output nodes (the unique output tokens). The activation functions used was SoftMax in the output layer. The model has in total 394,444 trainable parameters. We used Adam optimizer, and categorical\_cross\_entropy loss. 25% of the data were used for validation with 200 epochs. We used LSTM layer because of its ability to capture long sequence dependencies which is of great help in language modeling. The results obtained are discussed in the next section.

V. RESULTS AND DISCUSSION

This section demonstrates the experimental work done to evaluate our model. We used Google Colab GPU using a high RAM station to train the model. Training the 200 epochs took more than 6 hours. Due to the very long time of training, we used the first 50 epochs to show the accuracy results that was 92.12. However, after training the model for the complete 200 epochs later, we obtained an accuracy rate of 96.83%. The accuracy was a good metric for testing the model predictions

because the padding length didn't use lots of zeros because most of the data were concentrated in that length. However, we also used loss as an evaluation metric. The Model loss for the first 50 epochs is recorded in the figures below (Fig. 4). It reached below 0.19 for training and below 0.22 for validation. After the end of the 200 epochs, the loss obtained was 0.074 and the validation was 0.080. It seems clearly that the model isn't overfitting and obtaining excellent results.

To test the model, we needed to convert the sentence into all its corresponding characters and feed them into the trained model and recollect the characters predicted to get the output correct sequence. The results are clearly outperforming previous methods and are a good potential for the use of sequence-to-sequence modeling for automatic text diacritization. We can see that the main reasons for obtaining these excellent results are due to the correct data cleaning/preprocessing using regex and ell the text preprocessing steps. In addition, the padding length does not allow the use of lots of zeros in the sequence the thing that made the data represented in an excellent manner. In addition, the size of the validation data along with the absence of correlation between the features (because of one-hot encoding) allowed the model not to be overfitting. One of the main strengths of this work also, was that we extracted the exact unique tokens of both input and output which allowed the sequence-to-sequence model to perform excellently and outperform preceding sequence-to-sequence models. In the most recent research in literature RNNs and LSTMs were used but without the text preprocessing steps and size settings, they failed in obtaining high results using sequence-to-sequence models due to their inability to extract the exact unique tokens of diacritics and Arabic language.

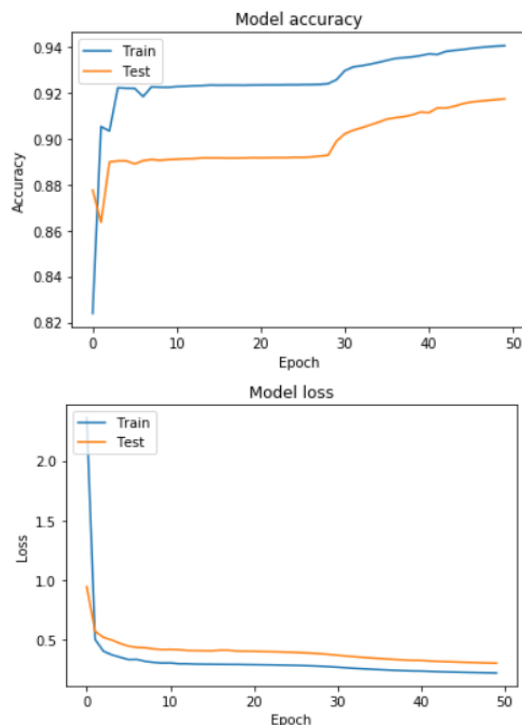


Fig. 4. The Accuracy/Loss vs epochs for the model training for the first 50 epochs.

## VI. CONCLUSION AND FUTURE WORK

This work achieved a high accuracy rate and low loss rate that outperforms results found in literature using sequence-to-sequence model to diacritize Arabic text. The main strength points in this work compared to related work in literature are:

- 1) The coverage of all Arabic diacritics rather than focusing on certain diacritics.
- 2) Diacritizing the whole word rather than diacritizing the end of it, and
- 3) The good text preprocessing and the extraction of unique tokens before using the LSTMs to diacritize the undiacritized text.

As a future work, we are planning to gather more data for further evaluation of our proposed model. We are also planning for using transfer learning to evaluate different deep learning models applied for the same problem.

## REFERENCES

- [1] Ahmed Emam, "Arabic Sentiment Analysis: A Survey", International Journal of Advanced Computer Science and Applications, December 2015.
- [2] Jezia Zakraoui, Saleh Moutaz, Somaya Al-ma'adeed, and Jihad Mohamad Aljaam "Arabic Machine Translation: A Survey With Challenges and Future Directions", IEEE Access, PP(99), December 2021.
- [3] Ahmed Abdelali, Hamdy Mubarak, Younes Samih, Sabit Hassan, and Kareem Darwish "Arabic Dialect Identification in the Wild", Arabic natural language processing tools, May 2020.
- [4] Chadi Helwe and Shady Elbassuoni, "Arabic named entity recognition via deep co-learning", Artificial Intelligence Review Journal, Springer, June 2019.
- [5] Mariam Biltawi, Sara Tedmori, and A. Awajan, "Arabic Question Answering Systems: Gap Analysis", IEEE Access, 2021.
- [6] Samira Lagrini and Redjimi Mohamed, "A New Approach for Arabic Text Summarization", The The Fourth International Conference on Natural Language and Speech Processing (ICNLSP), February 2021.
- [7] Aly Farghaly and Khaled Shaalan, "Arabic Natural Language Processing: Challenges and Solutions", ACM Transactions on Asian Language Information Processing, Vol. 8, No. 4, Article 14, Pub. date: December 2009.
- [8] Nizar Y. Habash, "Introduction to Arabic natural language processing", Synthesis Lectures on Human Language Technologies, 2010.
- [9] Abdulhadi Shoufan and Sumaya Alameri, "Natural Language Processing for Dialectal Arabic: A Survey", Proceedings of the Second Workshop on Arabic Natural Language Processing Conference, 2015.
- [10] Diab, M., Ghoneim, M., & Habash, N. "Arabic diacritization in the context of statistical machine translation", In Proceedings of MT-Summit. In Proceeding of the MT-Summit, Copenhagen, Denmark. 2007.
- [11] Abo Bakr H. M., Shaalan K., and Ziedan I., "A Statistical Method for Adding Case Ending Diacritics for Arabic Text", The Eighth Conference on Language Engineering, ESOLEC'2008, Page 225--234, Cairo, Egypt, December 17--18 2008.
- [12] Bouamor, H., Zaghouni, W., Diab, M., Obeid, O., Oflazer, K., Ghoneim, M., & Hawwari, "On Arabic Multi-Genre Corpus Diacritization", Qatar Foundation Annual Research Conference Proceedings, Issue 1 (Vol. 2016, No. 1, p. ICTPP2921), Hamad bin Khalifa University Press (HBKU Press), 2016.
- [13] Darwish, K., Mubarak, H., & Abdelali, A. "Arabic diacritization: Stats, rules, and hacks", Proceedings of the Third Arabic Natural Language Processing Workshop (pp. 9-17), 2017.

- [14] Fashwan, A., and Alansary, S., "SHAKKIL: an automatic diacritization system for modern standard Arabic texts". Proceedings of the Third Arabic Natural Language Processing Workshop (pp. 84-93), 2017.
- [15] Alosaimy, Abdulrahman, and Eric Atwell. "Diacritization of a Highly Cited Text: A Classical Arabic Book as a Case.", 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR). IEEE, 2018.
- [16] Abdelali, A., Attia, M., Samih, Y., Darwish, K., & Mubarak, H., "Diacritization of maghrebi arabic sub-dialects", arXiv preprint arXiv:1810.06619, 2018.
- [17] Fadel, A., Tuffaha, I., Al-Jawarneh, B., & Al-Ayyoub, M., "Neural Arabic Text Diacritization: State of the Art Results and a Novel Approach for Machine Translation", arXiv preprint arXiv:1911.03531, 2019.
- [18] M. Madhfar and A. M. Qamar, "Effective Deep Learning Models for Automatic Diacritization of Arabic Text", IEEE Access Journal, 2021.
- [19] Batool Abuali and Mohamad-Bassam Kurdy, "Full Diacritization of the Arabic Text to Improve Screen Readers for the Visually Impaired", Journal of Advances in Human-Computer Interaction, 2022.
- [20] Brian Thompson and Ali Alshehri, "Improving Arabic Diacritization by Learning to Diacritize and Translate", Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022), pages 11 – 21, May 2022.
- [21] Mona Diab, Nizar Habash, Owen Rambow, and Ryan Roth, "Ldc arabic treebanks and associated corpora: Data divisions manual", arXiv preprint arXiv:1309.5652., 2013.

# An Investigation of Cybersecurity Issues of Remote Work during the COVID-19 Pandemic in Saudi Arabia

Gaseb N Alotibi<sup>1</sup>, Abdulwahid Al Abdulwahid<sup>2</sup>

Faculty of Computer Science and Information Technology-University of Tabuk, Tabuk, Saudi Arabia<sup>1</sup>

Department of Computer and Information Technology, Jubail Industrial College, Saudi Arabia<sup>2</sup>

Royal Commission for Jubail and Yanbu, Jubail Industrial City, Saudi Arabia<sup>2</sup>

**Abstract**—COVID-19 pandemic has dramatically changed the public life style as well as the daily work activities across the world. This indeed has led both public and private sectors to attempt to adapt thereby shifting to remote work and adopting and enabling new technologies and online services in order to sustain their businesses while saving people lives. Unfortunately, a decent number of those endeavors have been undertaken unwarily in a hurry without taking the due diligence of all relevant aspects including cybersecurity and privacy. This survey aims at exploring the current state of the practice during the Covid-19 pandemic lockdown and the revolving challenges of using and publishing online services in Saudi Arabia. It also investigates the needs for investment in the cybersecurity field which would increase the trust on and reliability of them; and thus encouraging organizations to move confidently towards the real digital transformation.

**Keywords**—Cybersecurity issues; investigative survey; remote work; COVID-19 Pandemic; Saudi Arabia

## I. INTRODUCTION

COVID-19 pandemic has changed the life style of people in many aspects including social and work perspective. Indeed, the reliance on the Internet has become one of the daily life activities since the internet services have been used in most of the world countries. These activities include a variety of fields such as social network, entertainment, shopping, studying, and working. Therefore, about 4 billion users access the Internet, spending an average of 170 minutes daily – the majority of which was on social media [1]. However, the COVID-19 lockdown has forced the countries and their organizations in public/private sectors to empower remote working to keep their organizations running. Remote work is the means that allow employees to perform some or all of the work from home or any location other than their work site.

According to the statistical survey conducted by [2] to measure the remote system experience of epileptologists before and after the pandemic, they found that those who experience remote system have dramatically increased from about 63% to 86% within few months. In the UK, the remote working hits about 38% of the employees during the lockdown raising from 21% in 2019 [3]. This unforeseen noticeable changes in the policies and procedures of working that permeate over countries have made some challenges. These issues might be related to different aspects such as place, technology, and

security. The Forbes survey of remote workers showed that 72% of employees are not working from a dedicated office place [4]. In addition, the majority of them (56%) has faced a difficulty to bring their work equipment to their homes. In the report published by Global advisory and accounting networks [5], cyber security concerns were raised thereby showing that nearly 65% of organizations announced they have been either breached or exposed to cyberattacks during remote work. On the other hand, about 13% of the organizations were not prepared at all, compared to 45% somewhat prepared and 42% claimed they have prepared very well for the transformation in the work manner. Consequently, the damage of their income was incredibly hard in the majority of organizations across the world. According to the Office for National Statistics (ONS), the UK economy is still not close to the pre-corona virus output levels [6]. The report anticipated that the damage of the lockdown just in April 2020 might hit 15 billion to 20 billion GBP. Globally, the crisis could lead over 2020 and 2021 to an overall loss of around 9 trillion USD which is greater than two countries' economies such as Japan and Germany [7].

The cyber security aspect has been considered as main pivotal pillar in the remote work status. The struggle becomes clearer with the technology and security in response to the quick transfer in the work environment and probably paradigm. Indeed, the number of cyber attacks has dramatically increased to those organizations whilst processing information about the pandemic. For instance, during the pandemic crises, the World Health Organization (WHO) has mentioned that the number of cyber attacks to its staff and emails have surged more than five times compared with the same period a year before. Consequently, it is moving to a more secure authentication system after a leakage happened in its system [8].

Generally, the studies about the remote work have focused on the services that can be remotely provided. For example, [9] emphasized on some of the security challenges that need to be considered in each level of the cloud to secure the remote access to the site such as implementing and managing Identity and Access Management (IAM) systems. However, the pandemic incidence has forced organizations to think for a reasonable manner that can help in this partial or complete shift in their daily work. The author in [10] carried out a study focused on using PKI certificate and mobile device

management platform to secure the organization system during this accelerating change.

In this work, we investigate the challenges that have faced organizations within Saudi Arabia during moving to the remote work. In addition, it is going to concentrate upon the cybersecurity aspect and to what extent it might be limiting and/or enabling remote work transformation. The subsequent section explains the design methodology of the conducted survey. The Results Analysis Section presents a detailed illustration of the responses received and then followed by exploring and discussing them and succeeded by some drawn conclusions with a set of relevant future directions.

## II. METHODOLOGY

The survey was designed to explore and assess the maturity level of the organizations run in Saudi Arabia and its surroundings when they have moved to the remote work and challenges they have faced. Furthermore, it sought to explore the encountered information security breaches and utilized techniques during the COVID-19 pandemic. This was to answer the following research-related questions:

- What is the current state of the practice in the Saudi public and private sectors during the COVID-19 pandemic lockdown in relation to the growth of using and publishing online services and the revolving challenges and issues?
- Are cybersecurity concerns one of the vital issue(s) that may delay the digital transformation to online services?

A number of questions were devised and piloted with a number of specialized academic colleagues in order to obtain their perceptions on whether they are understandable and serve the purpose. After a few revisions based on the received inputs, the final version contained 28 questions and divided into four sections and structured as follows:

1) *Personal demographic:* Exploring the participants' demographic characteristics, related to gender, age, qualification, and job level.

2) *Organization information:* Establishing background of the organizations at which the participants work, in term of the type, field, size and location.

3) *Exploration of the current state of the practice:* Understanding technologies that were utilized into moving to remote work and during it, in addition to what extent online services have met the demand and requirements of stakeholders.

4) *Investigation of cyber security related issues:* Studying the reason(s) that has/have hindered or may hinder the growth of/ transformation to the remote work as well as security techniques that were used during it, their vulnerabilities and potential solutions.

The survey was set to be conducted over the Internet via an online questionnaire hosted by survey monkey website. Public users were targeted with three conditions: they are 18 years and above as well as were employees and worked in any means during the COVID-19 pandemic in Saudi Arabia. Those were

recruited via e-mail besides other social media, such as WhatsApp, Twitter and LinkedIn from acquaintances and professional societies and groups.

## III. RESULTS AND ANALYSIS

A total of 300 participants completed the survey over a period of eight weeks, during which the survey was active. Table I depicts that the majority of them (49%) belonged to the age group of 30-39 years, followed by 26% from the age group of 40-49 years. A huge gap between the proportionality of male and female participants was observed. About 89% of the partakers were male, reflecting the unequal participation of the different genders. However, as the investigated aspects of this research related to the organizations' employees, it is more likely to be common practices regardless of the genders. With respect to educational qualifications, the majority of the participants held bachelor's degree, followed by Master's, and Ph.D. – indicating high literacy levels of them and hence high probable informed responses.

Moreover, the majority of the participants assumed high job positions including mid-senior (41%), director (21%), and executive (7%) – showing good work experience and exposure with various areas of business operations. Accordingly, most partakers were working on the area of technology (30%), followed by 27% in education sector, and 12% in industrial field.

Fig. 1 shows that the participants were from diverse fields of work, representing more realistic and reflecting results.

TABLE I. SUMMARY OF PARTICIPANTS' DEMOGRAPHIC CHARACTERISTICS

Demographic Factors	Characteristics	Relative Frequency
Age	18-29	16.79%
	30-39	49.38%
	40-49	25.68%
	50 and above	8.15%
Gender	Male	89.38%
	Female	10.62%
Qualification	High School/ Associate Degree/ Diploma/	6.91%
	Bachelor	39.75%
	Masters	30.12%
	PhD	23.21%
Job Position	Entry Level	11.60%
	Associate	19.01%
	Mid-Senior	40.99%
	Director	21.23%
	Executive	7.16%

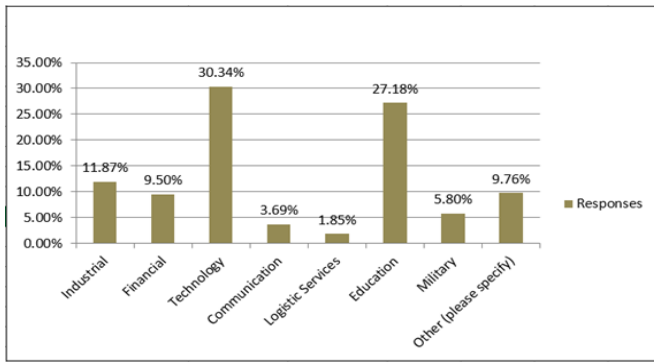


Fig. 1. Distribution of participants fields of work.

In addition, two thirds of the participants worked in public enterprises, while a third worked in private enterprises, indicating the distribution of participants in the two main sectors of the economy. In terms of the size of participants' organizations, 53% of the participants' organizations employed a thousand employees or more; 21% had between 100 and 999; and 19% had less. Thus, the size of the participants' organizations reflects the consideration of different small, medium, and large scale organizations, thus improving the analysis coverage of the results of this study.

Focusing on the strong technical infrastructure of the organizations that meets the needs of participants, it was identified that more than 77% of the participants stated that their needs were met. However, considerable number of participants (18%) were neutral, and about 5% stated that their needs were not met as illustrated in Fig. 2. Therefore, it can be a cause of concern for enabling and maintaining remote working conditions, suggesting a need for improvement in the technical infrastructure of the organizations.

Analyzing the use of devices in remote working conditions (as demonstrated in Fig. 3), it is identified that the majority of the participants relied on personal devices. 48% were using personal laptops, 17% personal desktop, besides 40% used personal smartphones/tablets for remote working. On the other hand, organizational laptops, desktops and smartphones/tablets were also being used by 48%, 23% and 10%, respectively. This increase in the use of personal devices for remote working may reflect the escalation in the security threats, as personal devices may be more prone to security attacks if proper security configuration is not maintained. Additionally, the diverse platforms usage indicates that most of the current remote workers probably own/use many digital devices with different operating systems and configurations –emphasizing the need to consider universal applicability as a crucial aspect in any proposed mechanism/solution.

Moreover, Fig. 4 represents the type of network connection utilized during remote working. The majority of the participants relied on private Wi-Fi connections (61%), indicating a more secured network compared to public Wi-Fi that was used by 6%. This may probably lead to potential security attacks that might affect the organizations and result in

huge losses. Whilst 43% utilized Fiber Optic and 12% relied on broadband services, 44% used mobile internet services. The latter indicates the possibility effect on the work flow due to mobile low speed or loss of connection. In this context, only 22% of the participants strongly agreed and 49% agreed that their internet service provider has met their expectation in terms of speed and connectivity. In addition, almost 13% of the participants stated that their internet service provider did not meet their expectations, while 17% were neutral. These results indicate that the internet and communications infrastructure was not fully upgraded/reliable in order to meet the requirements of remote working system.

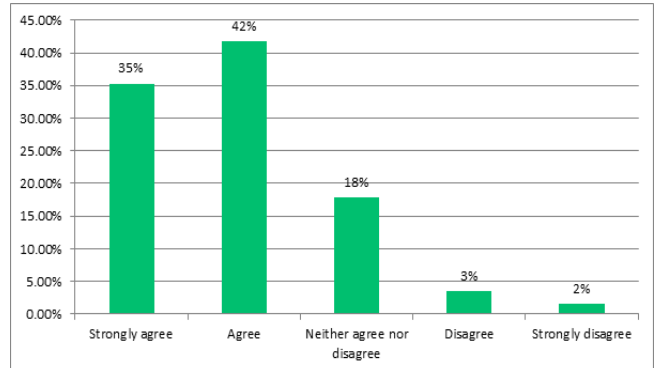


Fig. 2. The extent to which participants' organization meet their needs.

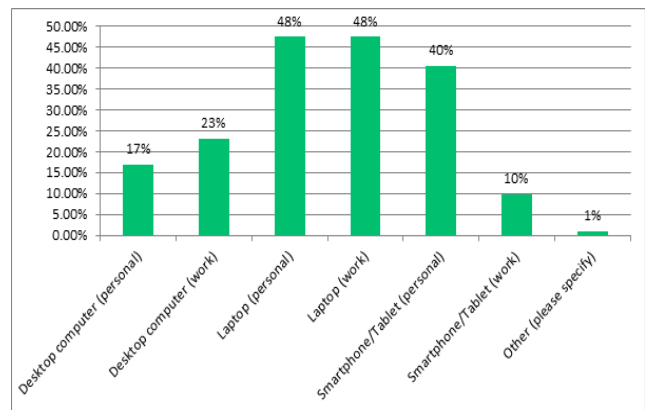


Fig. 3. Usage of digital devices during remote working.

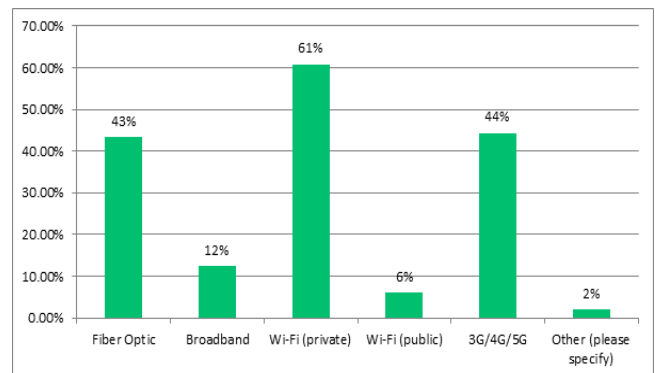


Fig. 4. Types of internet connections in use.

Considering the cloud services usage illustrated in Fig. 5, only 7% of the participants indicated that they do not use cloud services. Corporate cloud email services (60%) were the most preferred option by the participants, followed by the generic cloud file sharing services (34%) such as Google Drive, One Drive, Dropbox, iCloud, Nextcloud. As more than 90% of the participants used cloud computing services, it is likely that sensitive information may be accessed and updated during the remote working, leading to potential rise on privacy concerns. This, in addition to considering the diverse devices and network connections (from Fig. 3 and Fig. 4) used by the participants, cloud services may be more effective as they offer broad spectrum of universality and acceptability.

Regarding the security practices during remote working, Fig. 6 shows that only 5% of the participants did not use any authentication or security methods. Password/PIN/Pattern is the most commonly used authentication method adopted by the majority of participants (52%), followed by using VPN (39%) for remote working, and OTP (28%) for remote authentication. Face recognition and hardware token are other practices used by few participants. These results indicate that participants have adopted a range of security methods reflecting the use of multiple or a variety of security methods (44%) which adds an additional security layer for various activities during the remote working. Such practices improve security and privacy, thereby avoiding potential online security threats or loss of data.

Analyzing the impact of lockdown on the daily work of the respondents illustrated in Fig. 7, it can be identified that half of them stated that their daily work was affected, with 31% not affected and 19% was neutral. The results clearly indicate that lockdown has severely impacted their daily work routine while working remotely, indicating a greater influence of external factors (not internal to organization) on the daily work. Lockdown has not only affected the employees, but also the organizations, and their work culture. Most of the organizations shifted to remote working model in a far larger way than ever before. In this context, 76% of the participants stated that home Lockdown (Curfew) has encouraged or pushed their organizations to publish or enable more online services; while only 5% stated that it had no impact on their organizations in enabling more online services. Therefore, it can be inferred that there is a significant impact of lockdown on the organizational work paradigm, steeping more towards remote work.

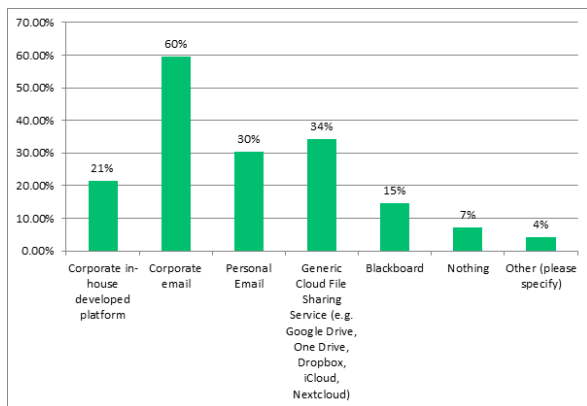


Fig. 5. Types of cloud file sharing services.

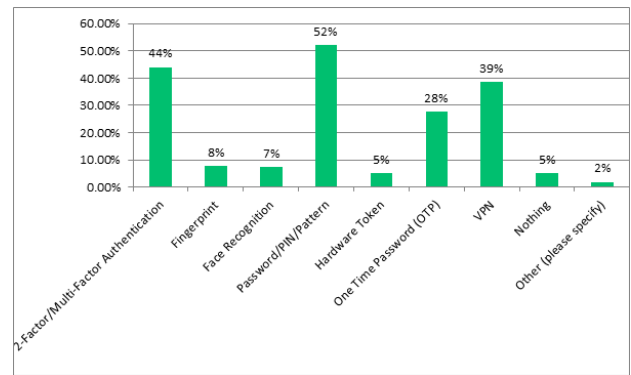


Fig. 6. Types of security / authentication practices.

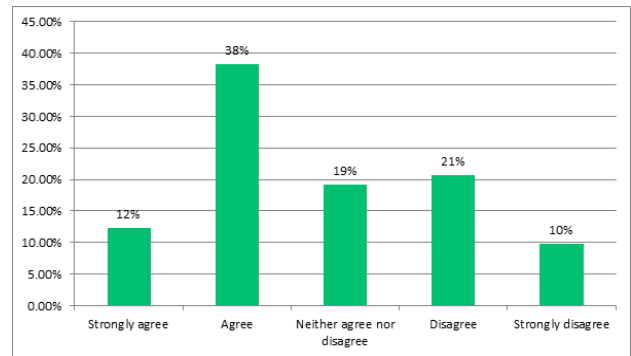


Fig. 7. Impact of lockdown on the daily work of the respondents.

Due to the sudden change in work culture, it is common that there could be many issues that might have to be dealt with by employees and organizations in shifting towards complete online services. In this context, Fig. 8 demonstrates the various challenges encountered by the participants and their colleagues during the remote working from home. The lack of high speed or quality connectivity infrastructure was the most common challenge identified by many participants (37%). These findings can be correlated with the earlier mentioned low satisfaction about their ISPs. Thus, it is evident that the lack of high-speed internet infrastructure in some areas of Saudi Arabia [11] and other states in the Middle East [12] was one of the major challenges affecting the remote working culture in the organizations. In addition, other major challenges included the lack of appropriate software (27%), the lack of appropriate hardware (23%), and the lack of appropriate skills/awareness (23%). These results indicate the lack of readiness of the employees and organizations in handling the unexpected change. However, it is unsurprising to note that 26% of the participants mentioned that they could not identify any challenge related to remote working during the lockdown, indicating an effective change management practices adopted by few participants and organizations during the lockdown.

Considering these sudden changes in the work culture, it is essential to compare the offline and online practices in order to improve the online services. Therefore, participants were asked to compare the organizational outcomes of online and offline services of their organizations. It is interesting to note that 62% of the participants reflected their opinion that online services used by their organizations delivered better outcomes than their counterpart offline services. Only 6% of the participants

indicated that offline services delivered better outcomes than online services. The results signified that adopting the online services and remote working culture resulted in better outcomes compared to offline practices - implying the advantages of remote working. Aspects such as flexibility, convenience, saving travel time and transport expenses might probably have contributed to the improved work efficiency of the participants while remotely working from home during the lockdown.

Fig. 9 indicates that Zoom was identified to be the most commonly used platform, followed by Cisco Webex, Microsoft teams, Blackboard, Skype for Business, and Google Meet. Due to its easy-to-use feature with limited required resources or memory, applications such as Zoom might have been preferred over other commercial applications such as Skype. Availability and adoption of various platforms for online meetings, as illustrated in Fig. 9, may improve the communication and interactivity among the various employees and entities within online environment resulting in improved process and operational efficiency leading to better outcomes. On the other hand, it was identified that password security method was the most commonly applied authentication technique by the majority of participants (58%) for joining online meetings. In addition, other authentication techniques used included OTP, VPN and multi-factor authentication. These security practices adopted online meetings can be correlated with the adopted general security practices (Fig. 6), reflecting similar practices.

Focusing on the various factors that may hinder the transformation from traditional work practices to online services, various important issues were identified by the participants. Cybersecurity, lack of infrastructure, and lack of trust were the three major issues identified for transforming traditional services into online services, as illustrated in Fig. 10. In addition, the lack of staff and users' skills and awareness were the succeeding key challenging factors identified in this study against moving towards digital teleworking. Accordingly, 82% of the participants held an opinion that cybersecurity may be adversely affected by the online services.

In this context, the participants were asked about the probable vulnerabilities that might occur due to the increase in online services and remote working. Fig. 11 presents that information/data leak was identified to be the most important vulnerability by the majority of the participants (76%). Furthermore, identity theft (55%), and unauthorized information access (41%) were the following most vulnerabilities identified.

In view with the use of different devices with different configurations, and the increased usage of personal devices illustrated in Fig. 3, the potential risk of security attacks would be higher. Accordingly, information leak/theft, unauthorized accesses are some common vulnerabilities that could occur, but can have huge impact depending on the type of data loss. Thus, the results indicate that there is a high probability of security attacks or vulnerabilities in remote working process. In reference to the probability of vulnerabilities occurrences, the

participants were asked if they had experienced any security incidents during working from home. Even though 73% indicated they did not experience any security incident, a non-negligible number of participants (16%) indicated that they did. This result can be in line with the level of cyber incidents in the Middle East, which may not only affect the remote working infrastructure, but also affect the organizations in a number of aspects. Moreover, 10% of the participants stated that they do not know if they have experienced any security incident. Unawareness about cyber incidents is another major issue, which may result in continuous exploitation by the attackers/hackers without being detected.

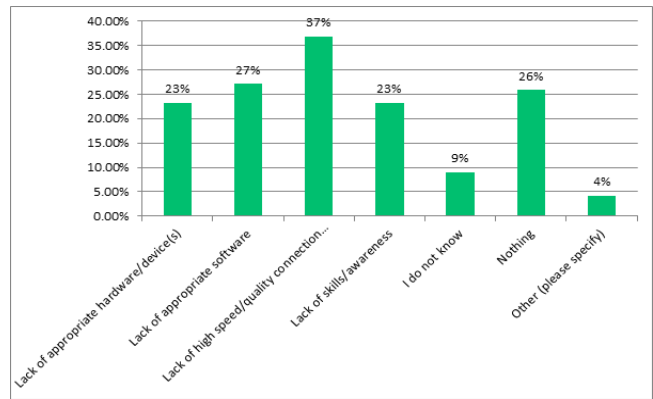


Fig. 8. Challenges faced by the participants in remote working during the lockdown.

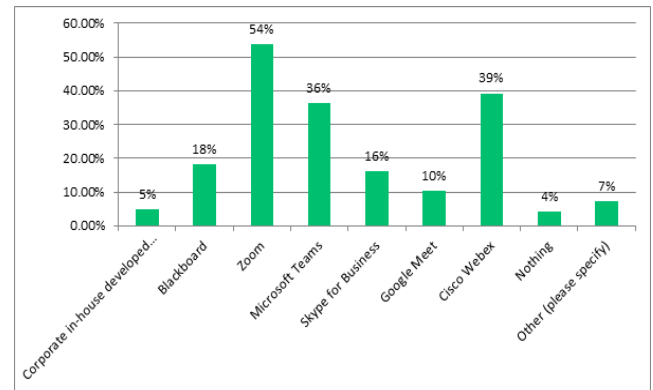


Fig. 9. Types of platforms used for online meetings.

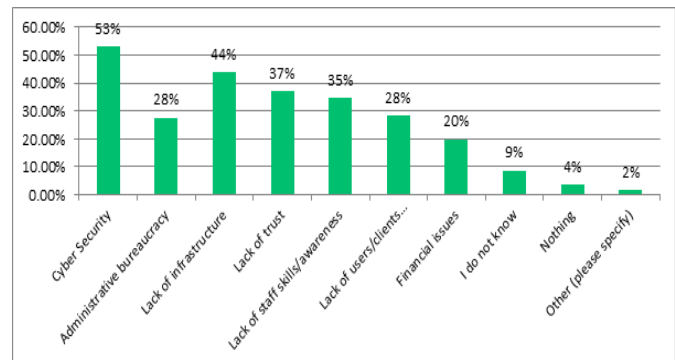


Fig. 10. Challenges in moving towards online services during the lockdown.



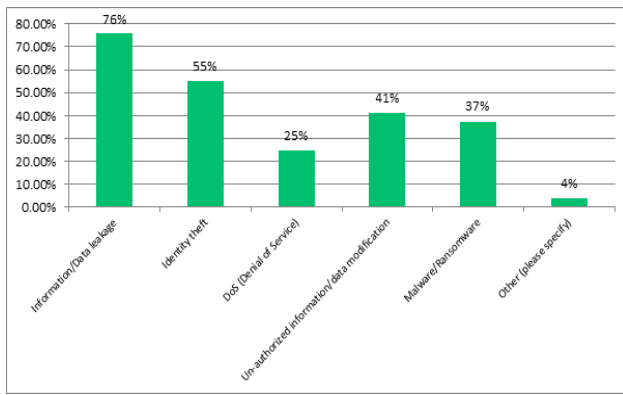


Fig. 11. Types of vulnerabilities that may occur during remote working process.

Considering the risk of security incidents during the remote working, it is important to analyze if the organizations are ready to invest in more effective cyber security solutions. Fig. 12 elucidates that over 80% of the participants strongly agreed and agreed that their organizations would invest more in cyber security solutions, indicating a growing importance for deploying effective security infrastructure in the process of transformation to online services. From the employees' perspective, it is essential to assess the impact of investments in cyber security solutions on trust and reliability in shifting to online services. In this context, over 87% of the respondents strongly agreed and agreed that the growth of investment by organizations in cyber security solutions will increase their trust and reliability in the growing online services and remote working.

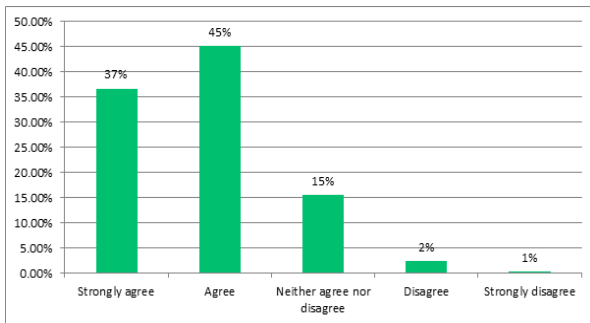


Fig. 12. Organizations' Readiness to invest more in cyber security solutions.

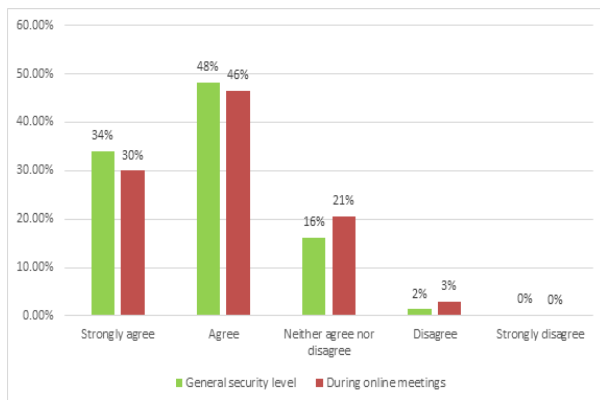


Fig. 13. Influence of transparent continuous authentication on enhancing the security levels.

With regard to the use of transparent and continuous authentication, participants' opinions were gathered whether it could enhance the security level in general and during online meetings specifically. Fig. 13 demonstrates similar results in both conditions, indicating a greater expectation and acceptance of transparent continuous authentication process.

#### IV. DISCUSSION

The respondents of the survey were from a diverse population with respect to age, qualification, job position, work field and enterprise sector and size, making it quite representative in relation to the investigated issues. Despite the overwhelming belief on the strong technical infrastructure of the organizations, there is an apparent increase in the use of personal devices for remote working, leading to potential surge of security threats due to the usual relaxed security practices while interacting with personal devices. Nevertheless, employing proper mobile device management (MDM) solutions would lead to more suited security measures and maintenance, hence alleviating cybersecurity vulnerabilities and attacks.

It is apparent that most of the current remote workers own/use a number of digital devices with different operating systems and configurations. This is coupled with the evident adopted several online meeting platforms and cloud services, emphasizing the need to consider universal applicability and interoperability as a crucial aspect in any offered security mechanism.

Given the fair range of applied security authentication practices, still the knowledge-based methods outweigh the others, rendering the remote working susceptible to its well-known shortcomings that affect the security level. This prevalence can be attributed to the fact that these approaches are the main provided ones by the majority of platforms, portals and services besides other low accurately configured and more intrusive alternatives such as the biometrics of fingerprint and facial recognition. Users' inclination to the ease of use may yield to apply the PIN/Pattern method, for example, at the expense of a further layer of protection that is introduced with biometrics as a single factor or an additional factor. However, implementing them in a usable and robust manner would lead to a resilient online working environment with less effect on the day to day tasks than what was declared. This can also be aided by better early preparedness in terms of appropriate devices, applications, quality internet connections, and staff skillsets and awareness in order to have improved smooth transition of processes and operations leading to efficient online remote or hybrid work paradigm.

It is observed that cybersecurity and the lack of infrastructure and trust are the major challenges of digital work transformation. This also can be seen by the respondents' perceptions regarding the wide range likely vulnerabilities during remote working including data leakage and unauthorized modification and identity theft. Therefore, an intelligent security solution that balance the higher protection and usability is required and can be provided by continuous and transparent authentication approach which gained high acceptability and expectancy by participants, stemming to a potential success. This is also supported by the anticipated

increase in investment in cybersecurity solutions by respondents' organizations.

## V. CONCLUSION

The survey findings were derived from a fair range of participants' backgrounds with regards to demographics, employments, as well as organization types, fields, sizes and somewhat locations. Exploring their state of the practice, it was found that a variety of technologies were adopted and operated into transforming to remote work and during it, in terms of devices, connectivity, Cloud File Sharing Services, and online meeting platforms.

Despite the acceptable level of satisfaction about the used and offered online services in meeting the demand and requirements of stakeholders, still a number of challenges were raised and needed to be tackled, such as cybersecurity and the lack of high speed/quality connection infrastructure, appropriate software/hardware/device(s), and technical skills/awareness.

With reference to issues that hindered or may hinder the growth of or the transformation to the remote work, cybersecurity had the highest concern including confidentiality, authenticity, integrity, malware and availability (in order). Moreover, the lack of infrastructure, trust, and financial support were of those matters to be considered with potential solutions for better remote work experience. Although many organizations have moved relatively fast to the remote work and to some extent fulfilled their tasks, many of them have faced difficulty to offer appropriate scale of online services. Therefore, more investment in cybersecurity techniques and employee training could be an incentive towards a more effective, efficient and smooth remote work environment.

## REFERENCES

- [1] Statista, "Internet usage worldwide – statistics & facts," available on <https://www.statista.com/topics/1145/internet-usage-worldwide/>, accessed on 29 November 2020, 2020.
- [2] M. Kuchenbuch, et al., "An accelerated shift in the use of remote systems in epilepsy due to the COVID-19 pandemic," *Epilepsy Behav.* ;112:107376, 2020.
- [3] L. Elliott, "Number working from home in UK rises after government U-turn," available on <https://www.theguardian.com/business/2020/oct/01/number-of-people-working-from-home-in-uk-rises-following-government-u-turn>, accessed on 29 November 2020, 2020.
- [4] C. Westfal, "Statistics show remote workers are frustrated, many still unprepared for working from home," available online, [www.forbes.com](http://www.forbes.com), accessed on 29 Jan 2022, 2020.
- [5] Hlb, "Global advisory and accounting networks," available on, <https://www.hlb.global/press-room/remote-working-leads-to-increases-in-cyber-attacks/>, accessed on 29 Mar 2021, 2020.
- [6] ONS, "Office for national statistic," available on, <https://www.ons.gov.uk/economy/grossdomesticproductgdp/articles/internationalcomparisonsofgdpduringthecoronaviruscovid19pandemic/2021-02-01>, accessed on 1 April 2021, 2021.
- [7] G. Gopinath, "The Great Lockdown: worst economic downturn since the Great Depression," available on, <https://blogs.imf.org/2020/04/14/the-great-lockdown-worst-economic-downturn-since-the-great-depression/>, accessed on 1 April 2021, 2020.
- [8] WHO, "WHO reports fivefold increase in cyber attacks, urges vigilance," available on, <https://www.who.int/news/item/23-04-2020-who-reports-fivefold-increase-in-cyber-attacks-urges-vigilance>, accessed on 12 Apr 2021, 2020.
- [9] F. Shahzad, "State-of-the-art survey on cloud computing security challenges, approaches and solutions," in *Procedia Computer Science*, 37, 357-362, <http://dx.doi.org/10.1016/j.procs.2014.08.053>, 2014.
- [10] B. Trzuppek, "PKI is key to securing a post-Covid remote workforce," in *Computer Fraud & Security*, vol. 2020, no. 10, pp. 11-13, October 2020.
- [11] H. Alshehri and F. Meziane, "Current state on internet growth and usage in Saudi Arabia and its ability to support e-commerce development," in *Journal of Advanced Management Science*, vol. 5, no. 2, pp. 127-132, 2017.
- [12] Houcheimi, "The key e-tail opportunities and challenges in the Lebanese e-commerce market," *Journal of Information System and Technology Management (JISTM)*, vol. 7, no. 26, pp. 13-31. 10.35631/JISTM.726002, 2022.

# Patent Text Classification based on Deep Learning and Vocabulary Network

Ran Li<sup>1</sup>, Wangke Yu<sup>2\*</sup>, Qianliang Huang<sup>3</sup>, Yuying Liu<sup>4</sup>

Intellectual Property Information Service Center & School of Management and Economics  
Jingdezhen Ceramic University, Jingdezhen, China<sup>1,3,4</sup>  
School of Information Engineering, Jingdezhen Ceramic University  
Jingdezhen, China<sup>2</sup>

**Abstract**—Patent documents are a special long text format, and traditional deep learning methods have insufficient feature extraction ability, which results in a weaker classification effect than ordinary text. Based on this, this paper constructs a text feature extraction method based on the lexical network, according to the inner relation between words and classification. Firstly, the inner relationship between words and classification was obtained from linear and probability dimensions and the lexical network were constructed. Secondly, the lexical network is fused with the features extracted from the deep learning model. Finally, the fusion features are trained in the original model to get the final classification result. This method is a classification enhancement method that can classify patent text alone or enhance the accuracy of various types of neural networks in patent text classification. Experimental results demonstrate that the accuracy of BERT combined with lexical network method is as high as 82.73%, and the accuracy of lexical network method combined with CNN and LSTM is increased by 2.19% and 2.25% respectively. In addition, it was demonstrated that the lexical network feature extraction method accelerated the convergence speed of the model during training and improved the classification ability of the model in Chinese patent texts.

**Keywords**—Text classification; deep learning; network vocabulary; patent; feature extraction

## I. INTRODUCTION

As the main carrier of scientific and technological information, patents contain about 90% of global technical information. In 2021, the effective number of Chinese invention patents will be 3.597 million, facing a large number of patent applications, patent text review takes up a lot of time for patent examiners. Patent classification is an important part of the patent review, which can reduce the time and cost of subsequent retrieval and management of patent documents. At present, the main patent classification used in China is the International Patent Classification [1]. IPC classification has five levels, belonging to a multi-level and multi-label classification system. At present, China mainly uses manual methods to classify patents. For example, after receiving patent application documents, patent examiners need rich professional knowledge to accurately classify patents, making subsequent patent texts easy to review and retrieve. In the face of massive patent texts, intelligent and effective implementation of

automatic classification of Chinese patent texts has become the focus of the industry and academia.

Automatic text classification is one of the important fields in NLP (Natural Language Processing). In recent years, there has been much research on text classification using machine learning and deep learning, Early text classification models generally refer to the support vector machine[2][3], Bayesian classifier[4][5], k-Nearest Neighbor algorithm[6][7] and other machine learning[8] methods for text classification. However, text classification of machine learning methods relies on manually constructed features and ignores the context between words and sentences, which makes it difficult to improve the accuracy of such methods. After Hinton[9] and others first proposed the concept of deep learning in 2006, scholars proved that deep learning technology has good performance in text classification. Since then, in the field of text classification, deep learning has gradually replaced machine learning as the mainstream research direction.

Therefore, on the basis of previous studies, based on the analysis of a large number of patent data, this paper proposes a deep learning feature extraction method based on a vocabulary network, which improves the accuracy of deep learning on patent text classification through feature fusion between the extracted features and the deep learning model. The practice shows that this method combined with deep learning has a certain improvement in Chinese patent text classification, which confirms the contribution of this paper in the following aspects.

1) In this study, a method of combining deep learning with vocabulary network feature extraction is proposed theoretically.

2) It is proved in practice that this method combined with other deep learning models can speed up the model convergence and improve the accuracy of model classification.

The rest of this paper is organized as follows: In Section II, we introduced the relevant work and our contributions. In Section III, we introduce our research framework. In Section IV, we selected text data to verify and illustrate the relevant parameters of this model. Finally, the conclusion and further work are given.

---

\*Corresponding Author

## II. RELATED WORK

Automatic patent text classification is a technical means to use computers to classify and mark the text according to a specific label system, which is convenient for the patent document management and text search. This paper mainly solves the task of automatic patent text classification through deep learning. Therefore, this section mainly introduces the current research status of scholars using deep learning methods to classify text.

Deep learning text classification is mainly represented by a neural network model, attention mechanism and pre-training model.

1) The text classification method based on the neural network model presents the text in low dimension by word embedding and achieves the classification effect by using an encoder. Representative models include convolutional neural network (CNN)[10], recurrent neural network (RNN)[11] and graphical neural network (GCN)[12], Among them, CNN can acquire local relations[13], RNN uses a recurrent neural network to have the ability to learn long-distance texts[14], GCN can effectively integrate the features of references.

2) Compared with the neural network model, the attention mechanism can focus on a part of the information related to the classification task while ignoring the irrelevant part when processing tasks, In 2014, Bahdanau et al[15].Used the attention mechanism to perform machine translation tasks for the first time, making great progress in machine translation. In text classification tasks, the attention mechanism is often combined with the neural network model, She et al[16]. Proposed a text classification algorithm based on the hybrid CNN-LSTM model, used the feature vector output from CNN as the input of LSTM and used Softmax classifier for classification, which can effectively improve the accuracy of text classification. Yang et al[17]. Based on the combination of convolutional neural networks and long-term and short-term memory, are used to capture local features to enrich feature information and use weight values to adjust the enhancement intensity.

3) The pre-training model is represented by the BERT model, and the improvement of model classification results mainly depends on the adjustment process of the model[18]. BERT has a strong sentence and word expression ability and has an excellent performance in NLP tasks at the word level[19]. Lu et al[20]. Proposed the VGCN-BERT model, which combines BERT's capabilities with vocabulary graph convolutional network (VGCN). The different layers of BERT interact to enable them to interact and jointly build the final classification representation. Zheng et al[21]. Proposed a BERT-CNN model for text classification. By adding CNN to the task-specific layer of the BERT model, you can obtain information about important fragments in the text.

Some of the above methods have been successfully applied to patent text classification tasks, and have achieved some results. As a branch of special long texts in text classification, Chinese patent texts are difficult to improve the accuracy of

classification. Many scholars have carried out a lot of research on this. For example, Wen et al. [22] proposed a multi-level patent text classification model, called ALBERT-BiGRU, which combines ALBERT and bidirectional gated loop units. The dynamic word vector pre-trained by ALBERT is used to replace the traditional static word vector, and the BiGRU neural network model is used for training, which preserves the semantic association between long-distance words in the patent text to the maximum extent. Roudsari et al[23]. Studied the effect of applying the DistilBERT pre-training model and fine-tuned it to complete the important task of multi-label patent classification. To sum up, domestic and foreign scholars' patent text classification methods based on deep learning mainly focus on the fusion application of deep learning models, the adjustment of pre-training models or the introduction of attention mechanisms to extract relevant features. Although these methods can achieve the goal of improving the classification effect, they are all cumulative applications of existing methods, which not only affect the model classification efficiency but also difficult to improve the patent classification effect.

In previous studies, the focus has been on improving the models, while ignoring the specific structure of the data itself, making it difficult to improve the accuracy of patent text classification. Based on this, a generic feature extraction method is considered for patent text feature extraction to improve the accuracy of automatic patent text classification, mainly considering the intrinsic connection between words and classification as the basis for feature extraction, and introducing the extracted features into the deep learning model to fuse with the features extracted by the model to improve the model classification effect.

## III. PROPOSED WORK

Through the analysis of previous text classification studies, it is found that LSTM has more advantages than CNN in text classification because of its contextual characteristics[24], which has also been confirmed in many long text classification studies. However, in the classification of Chinese patent texts, there is an exception in that the accuracy of CNN is equal to or slightly higher than LSTM. Through the analysis of patent data and experimental reproduction, it is found that patent documents belong to a special format in long texts, and are insensitive to the context, but sensitive to the words divided by social production practices such as industry and industry. Therefore, in the neural network feature extraction stage of deep learning, the parameters extracted by traditional feature extraction methods in patent texts have the low feature density required for classification, and the existence of multi-word synonyms in feature words increases the difficulty of classification, which makes the full connection layer learning feature inefficient and feature learning incomplete in deep learning, ultimately leading to low accuracy of Chinese patent text classification. In order to solve the problem that the feature extraction ability of Chinese patent text of neural network is insufficient, this paper proposes a feature extraction method based on a vocabulary network to strengthen the feature extraction ability of Chinese patent text of deep learning network, so as to improve the accuracy of Chinese patent text classification.

A. Deep Learning Model based on Vocabulary Network  
Feature Extraction

This paper proposes a feature extraction method of a deep learning model based on a vocabulary network. This method combines the extracted features with the extracted features of the original model and expands the features on the basis of retaining the features of the original model to improve the classification accuracy. Fig. 1 is a schematic diagram of the LNFE method of deep learning model fusion, which includes four parts: text vector representation, feature extraction, feature

fusion and text classification prediction. In the figure,  $P_i$  is patent data,  $W_i$  is patent text information after word segmentation, and  $F_i$  is the feature information extracted from text. In the training or prediction process, the patent text is first segmented and processed separately. In the deep learning network feature extraction, the data is extracted in the original feature extraction layer and LNFE, respectively, and then the extracted features are spliced and fused, and then the appropriate parameters are trained in the full connection layer to classify the subsequent data.

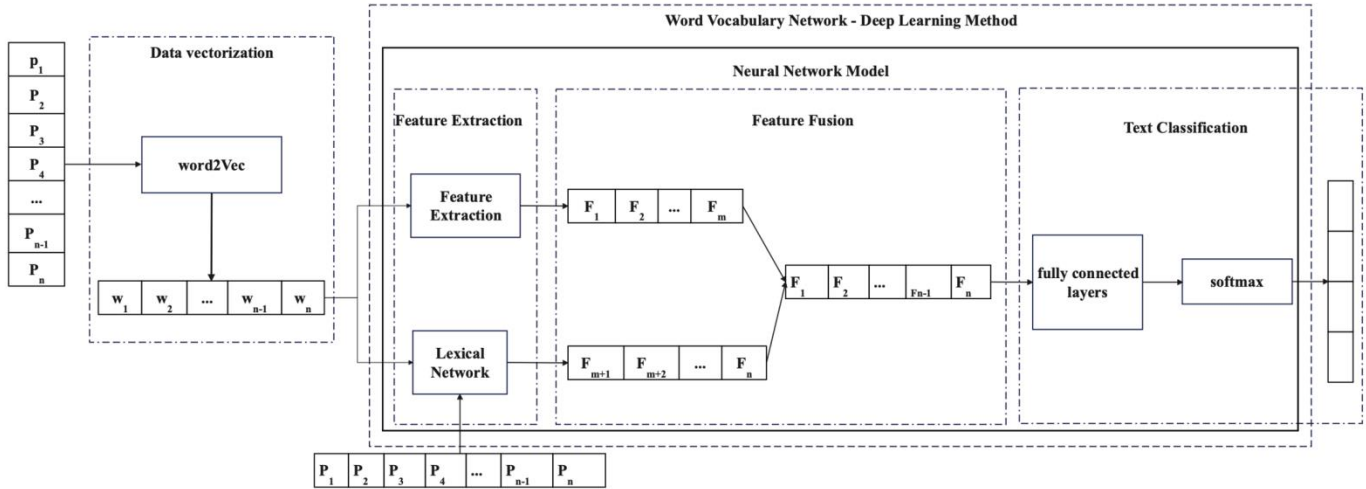


Fig. 1. Schematic diagram of vocabulary network and model fusion

B. Vocabulary Network Generation

The lexical network is a key step in feature extraction, which includes synonym clustering of words in the patent, the internal relationship between words and classification and other information. In the subsequent feature extraction process, the characteristic words of patent text can be accurately extracted using the lexical network as a reference. Fig. 2 is a schematic diagram of the generation of a vocabulary network.

First, the word vector (Word Embedding) representation of each word in all texts is obtained by unsupervised learning of Word2vec, and the word vector similarity is calculated by using the cosine formula of Formula 1 to cluster and merge synonyms in words, thus avoiding the problem of slow model convergence caused by synonyms in the feature extraction stage.

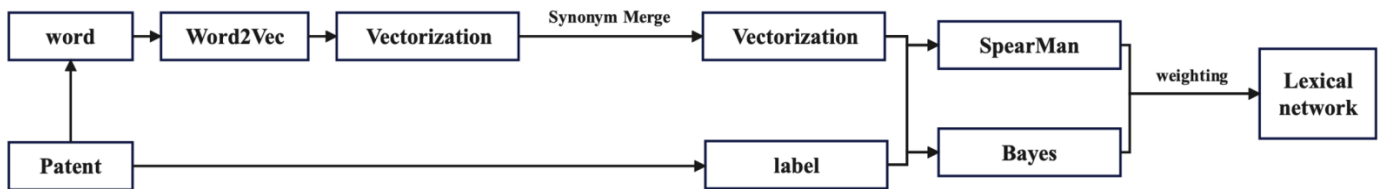


Fig. 2. Schematic diagram of vocabulary network generation

$$\text{Sim}(\vec{v}, \vec{w}) = \sum_i^N v_i w_i \quad (1)$$

Where  $\vec{v}$  and  $\vec{w}$  are word vectors trained by Word2vec respectively, Sim is the similarity between vector  $\vec{v}$  and  $\vec{w}$ ,  $v_i$  and  $w_i$  is the value of the vector  $\vec{v}$  and  $\vec{w}$  at different latitudes.

Each group of word vectors after merging has corresponding label labels. According to Formula 2 and Formula 3, the prior probabilities of words and classifications are obtained from the two dimensions of probability and linearity through Label information and word vector

information, and each word vector is given a classification feature weight by weighting.

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^n P(A|B_i)P(B_i)} \quad (3)$$

Where the  $\rho$  Is the correlation between word vector and classification,  $x_i$  is the frequency of word change in patent No.  $i$ ,  $\bar{x}$  appears in this category  $\bar{x}$  Average frequency of.  $y_i$  is  $x_i$

Classification of  $y$  in  $i$  word patent,  $\bar{y}$  is the average of the categories in the patents with  $x_i$  words. In Formula (3),  $P(B_i|A)$  represents the probability of being classified as  $B_i$  when the word  $A$  appears,  $P(A|B_i)$  is the probability of occurrence of  $A$  in  $B_i$  classification,  $P(B_i)$  is the probability of occurrence of  $B_i$  classification.

### C. Classification Model Design

Based on the characteristics of existing models, this paper selects CNN, LSTM and Bert models to combine with LNFE as the new deep learning model, and designs LNFE, CNN, LSTM and Bert classical models as the contrast to judge the classification effect of the new deep learning model.

- LNFE-CNN model

LNFE can use patent text feature extraction to form a vocabulary network. CNN cannot extract the context, semantic information of the patent text. The LNFE-CNN model combines the characteristics of a vocabulary network and convolutional neural network to improve the training speed. Fig. 3 is the schematic diagram of the LNFE-CNN model.

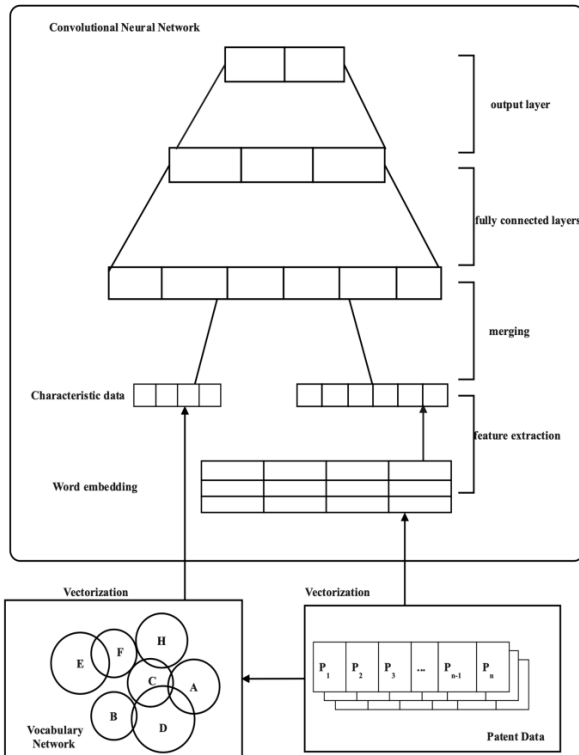


Fig. 3. Schematic diagram of the LNFE-CNN model

- LNFE-LSTM model

LNFE-LSTM model short-term memory is a special RNN network, which is mainly used to solve the problem of gradient disappearance and gradient explosion during long sequence training. LSTM mainly has three gate structures: forgetting gate, input gate and output gate. The forgetting gate, mainly decides to give up the information and leave important information. The input gate, mainly determines which information is added, and the output gate determines which

information can be output after judgment. The supplementary calculation formula of the LNFE-LSTM model is as follows:

Forgotten Gate

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

Input gate

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (6)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (7)$$

Output gate

$$O_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O) \quad (8)$$

$$h_t = O_t \cdot \tanh(C_t) \quad (9)$$

In this paper, we propose an LNFE-LSTM model, which can accurately extract patent features through LNFE, combine LSTM text classification to deal with long text classification problems, and has the characteristics of context semantic features.

- LNFE-Bert model

The Bert model perfectly considers the bidirectional semantic features, while retaining the maximum meaning of the text according to the context word order features. This paper accurately extracts the patented features through LNFE, combines Bert (Bidirectional Encoder Representations from Transformers) [25] with the bidirectional transformers encoder, trains the model through the masked language model and next sentence prediction, Have strong sentence expression ability. The supplementary calculation formula of the LNFE-Bert model is as follows:

$$(\theta, \theta_1, \theta_2) = L_1(\theta, \theta_1) + L_2(\theta, \theta_2) \quad (10)$$

$$L_1(\theta, \theta_1) = -\sum_{i=1}^M \log_p(m = m_i | \theta, \theta_1), m_i \in [1, 2, \dots, |V|] \quad (11)$$

$$L_2(\theta, \theta_2) = -\sum_{j=1}^M \log_p(n = n_j | \theta, \theta_2), n_j \in [IsNext, NotNext] \quad (12)$$

### D. Evaluation Indicators

In order to evaluate the classification effect of the LNFE method combined with deep learning, this score uses Precision, Recall and F1 values as evaluation indicators, and judges the feature extraction ability of the LNFE method by comparing various indicators of LNFE, CNN, LNFE CNN, LSTM, LNFE LSTM, Bert, and LNFE Bert models. Formula (13-15) is the calculation formula of the P value, R-value and F1. Generally, the higher the P value is, the higher the prediction accuracy is which proves that the stronger the prediction ability of the model is, the higher the R-value is, the better the model is, and the better the classification effect is. Since the single expression of the P value and R-value may lead to incomplete measurement standards due to the uneven distribution of data, F1 value can better evaluate the model by means of a weighted average.

$$P = \frac{\text{Predicted number of correct samples}}{\text{All predicted results}} \quad (13)$$

$$R = \frac{\text{Predicted number of correct samples}}{\text{All results in the sample}} \quad (14)$$

$$F_1 = \frac{2 \cdot P \cdot R}{P + R} \quad (15)$$

#### IV. MODEL EVALUATION AND RESULT ANALYSIS

##### A. Experimental Platform and Parameter Selection

The processor of the experimental platform used in this paper is Dual Core Intel Core i7, the memory is 8 GB (2133 MHz) LPDDR3, and the GPU is Nvidia 2080Ti. The operating environment is Python 3.9, the network framework is PyTorch, and the training models are CNN, LSTM, BERT, LNFE, LNFE-CNN, LNFE-LSTM, and LNFBERT. Among them, CNN, LSTM and BERT are general comparison networks in current patent classification research, and LNFE is a neural network that only uses the LNFE method as a feature extraction method. LNFE-CNN, LNFE-LSTM and LNFE-BERT are neural networks constructed by combining the extracted features of LNFE with the original features. Through analysis, it is found that the LNFE method has a high correlation with the classification in the first 40 words on average in the extracted feature times, so the LNFE method intercepts 40 words for training. In the training process, Table I is selected as the parameter of the neural network for training.

TABLE I. NEURAL NETWORK PARAMETER SETTING

net	learn rate	bach size	word size	word vector
CNN	5.00E-4	64	240	200
LSTM	5.00E-4	64	240	200
BERT	3.00E-5	32	240	200
LNFE	1.00E-3	32	40	240
LNFE-CNN	1.00E-3	64	240+40	200
LNFE-LSTM	1.00E-3	64	240+40	200
LNFE-BERT	5.00E-5	32	240+40	200

##### B. Data Source and Preprocessing

In this experiment, the "part" in the main IPC classification number is selected as the classification basis. The classification data is from the incoPat patent database. The obtained patent information includes the title, abstract, IPC main classification number, independent claims and other information. The uniform distribution of data in the learning features of the model is conducive to improving the learning effect of the model. Therefore, patent data are randomly selected from the eight IPC ministerial classifications of patents, and 125064 patents are screened through data denoising. In order to evaluate the effectiveness of the model, cross-validation method is adopted. The paper experiment extracts 80% of the collected data as the training set, 10% as the verification set, and 10% as the test set. The data distribution of the patent data set is shown in Fig. 4.

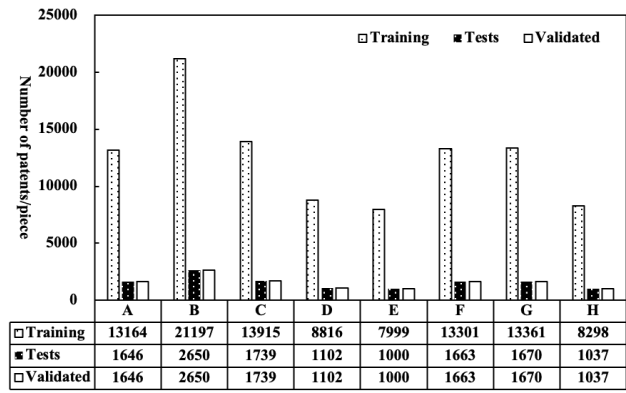


Fig. 4. Data distribution

##### C. Patent Text and Word Vector

The title and abstract of a patent are a summary of patent documents. Independent claims fully represent technical solutions and necessary technical features. This information can be used as a feature source of patent classification. To classify patents using deep learning, we first need to import the text to a word vector into the neural network, so we need to use Word2vec to convert Chinese patent text information into a word vector and then import it into the neural network for learning. This paper combines several neural networks, including Word2vec, CNN, LSTM, Bert, and so on. The vocabulary length of the training text is fixed during the training process. Table II shows the statistics of the vocabulary length of patent texts. The vocabulary queue of most patent texts is between 200 and 400, with an average vocabulary length of 241. In this paper, 240 words are used as the text cutting position, and the missing parts are filled with 0 vectors.

TABLE II. MODEL PARAMETER SETTINGS

Data	Sample length below 200 words	200 words - 400 words	Sample length above 400
Training	12309	62819	20192
Validation	712	13890	1201
Test	819	13210	1192

##### D. Vocabulary Network Generation

In the process of vocabulary network generation, 125064 patent documents generated 272141-word vectors, and 241524-word vectors remained after removing stop words and merging synonyms. By calculating the Spearman correlation and Bayes probability weighting of words and tags, 30080 words with strong correlation characteristics were obtained and classified, and their word network is shown in Fig. 5. Each point in the vocabulary network represents a word. The position of the point is the mapping of the high latitude of the word vector in the low dimensional. The shorter the distance between the points, the closer is the part of speech. The colors of the dots represent different parts of the patent classification related to the word. It can be found from the number of related words that a large number of words have a strong relationship with some categories. From the distribution of words, the meaning of words is also related to classification, which also confirms the above statement that "Chinese patent texts are sensitive to words classified by social production practices such as industries and industries".

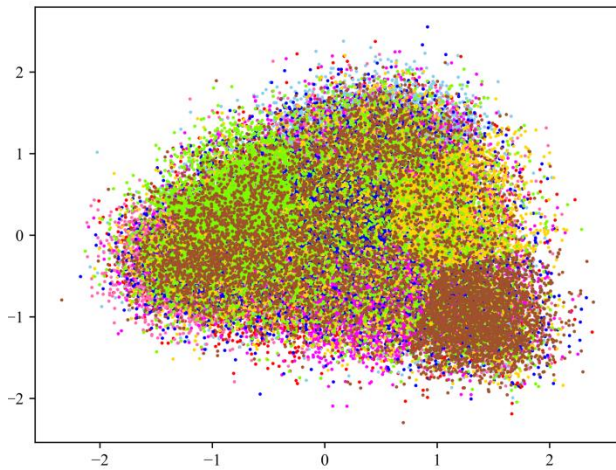


Fig. 5. Vocabulary network diagram

### E. Model Training

The deep learning model used for classification is usually trained at the full connection layer after feature extraction. Finally, the data are classified through SOFTMAX. After repeated training, the accuracy of classification is improved. In order to verify whether the LNFE method has the ability of

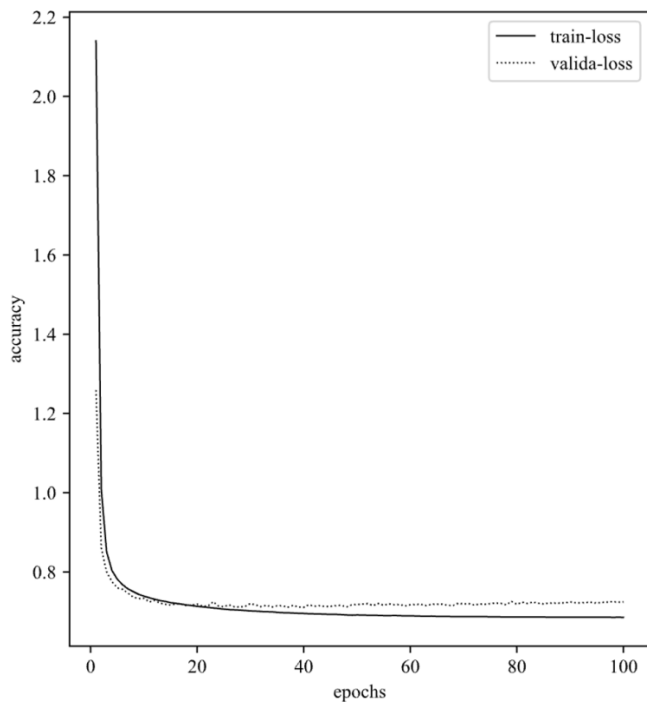


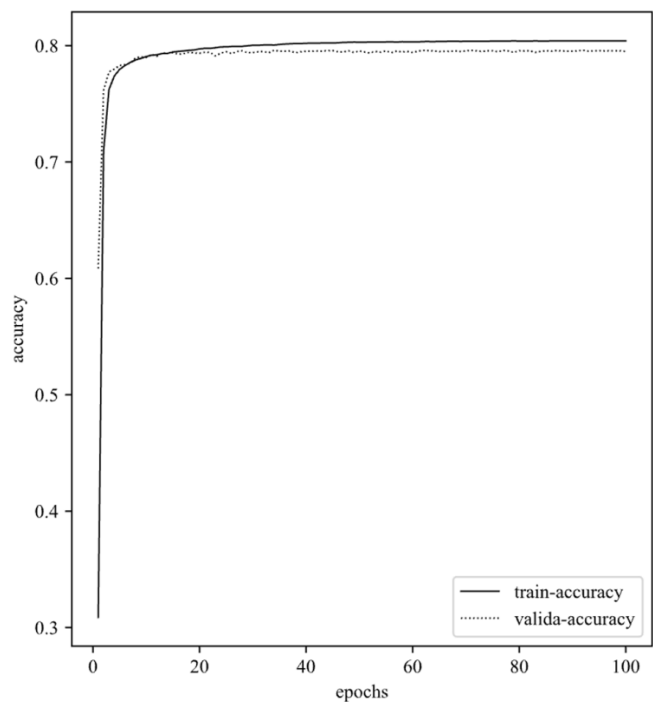
Fig. 6. LNFE training process

Word2vec+LNFE method is slightly higher than other networks; it is lower than traditional neural networks combined with the LNFE method. This is because the LNFE method only extracts features in probability and linear dimensions, while

feature extraction, the LNFE method is used as the feature extraction layer of the neural network to test the classification results. First, after removing the stop words and vectorization of the Chinese patent text, the similarity analysis of the word meaning with the words in the vocabulary network is conducted, and the words in the word segmentation results of the patent text are replaced with highly similar words in the vocabulary network to merge the multi-word consent words. According to the weighted results of words and classification, the word vector of 40 words with the highest weight is selected as the feature, which is sent to the network layer for training. The training process is shown in Fig. 6. After several iterations of training, the accuracy of the model can be found to improve rapidly. This result confirms that the LNFE method indeed extracts classified feature data.

### F. Accuracy Verification

After the LNFE method is confirmed to have the feature extraction capability, this section verifies the feature extraction effect of LNFE by comparing it with other networks. Table III shows the comparison results between the LNFE method and other networks. During the validation process, it is found that traditional neural networks have improved after combining the LNFE method, while Bert has limited improvement ability to Bert due to its excellent feature searching ability. Although the



traditional networks consider more dimensions. Because the LNFE method has the characteristic of being Embeddable for neural networks, it has practical value.



TABLE III. CHINESE PATENT TEXT CLASSIFICATION RESULTS

Number	Model	P/%	R/%	F1/%
1	Word2vec + LNFE	81.16	73.24	77.00
2	Word2vec + LSTM+ Attention	78.83	65.23	72.20
3	Word2vec + LNFE-LSTM + Attention	81.02	69.12	74.60
4	Word2vec + CNN	79.32	66.43	72.89
5	Word2vec + LNFE-CNN	81.57	74.51	77.88
6	Bert	82.31	76.13	79.10
7	LNFE-Bert	<b>82.73</b>	<b>78.31</b>	<b>80.27</b>

### G. Data Analysis

The classification results after CNN and Bert classify separately and combined LNFE method is shown in Fig. 7. Because the text features found by the LNFE method are more explicit, LNFE CNN has a faster convergence speed than CNN. At the same time, LNFE CNN may learn more text features than CNN, so the accuracy of LNFE CNN is slightly higher than CNN. The LNFE method has the prior probability of

words in the patent text so that LNFE Bert can find the complex functions of features and classification earlier than the Bert model in the training process, that is, the LNFE Bert model converges earlier than the Bert model. However, Bert's model pays more attention to the feature extraction of text. After a lot of training, the accuracy of the Bert model and LNFE Bert is almost the same. Therefore, it can be determined that the LNFE method can improve the learning ability and classification accuracy of traditional deep learning models.

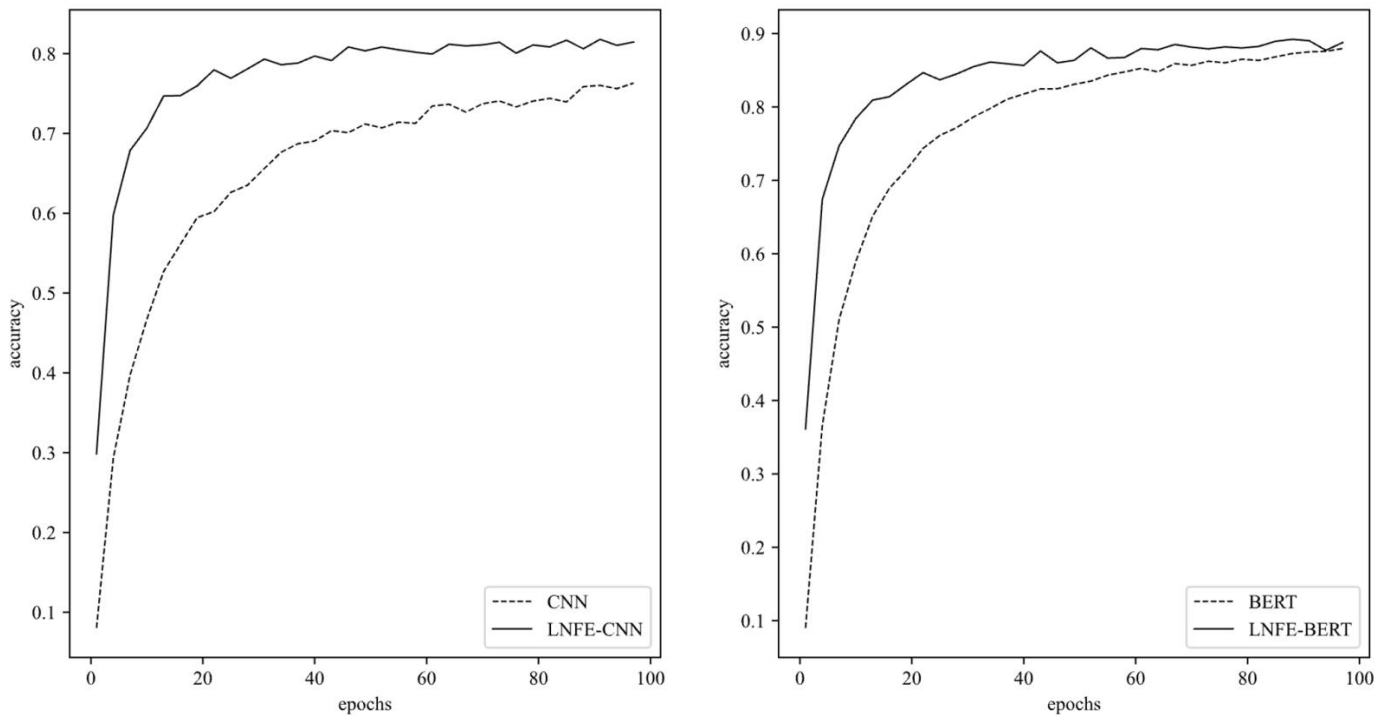


Fig. 7. LSTM and LNFE-LSTM training process chart

### V. CONCLUSION

When comparing many text classification materials, it is found that the accuracy of the classification results of Chinese patent texts is lower than that of other texts. Through repetition experiments and analysis and verification, it is concluded that patent documents are insensitive to context due to special formats, and the traditional deep learning model has insufficient ability to extract features of Chinese patent texts. These two reasons lead to slow convergence speed and low accuracy of the model. In order to solve this problem, this

paper proposes an LNFE feature extraction method, which uses Spearman and Bayes to obtain the prior probabilities of vocabulary and classification from the probability and linear dimensions respectively, constructs the network relationship between vocabulary and classification in a weighted way, and uses the features extracted from the vocabulary network to fuse with the features extracted from the deep learning model for classification. The method differs from the traditional method of improving the accuracy of the model by altering the neural network, by extracting the patent text features in depth through the prior probability of the text, thus allowing the combination

of various neural networks to enhance the performance and accuracy of the network in patent text classification. The experimental results demonstrate that the method speeds up the convergence of the model to a certain extent and improves the classification accuracy of the model. However, in this paper, only a ministerial lexical network is constructed to validate the method for Chinese patent text classification. In future work, we will try the method for more fine-grained patent classification in the case of feature extraction, and try the method in other text classification to verify whether the method is universal.

In future research, we will enrich the construction methods of patent text features based on how to make more subtle potential features of patent text, so as to improve the classification performance of the model. At the same time, it will also further explore the performance of the model applied to other tasks in the field, and consider how to deal with multi-classification problems to provide better help for researchers.

#### ACKNOWLEDGMENT

This article was supported by the Special Project for Scientific and Technological Cooperation of Jiangxi Province Grant Number [20212BDH80021].

#### REFERENCES

- [1] H. Sasaki and I. Sakata, "Identifying potential technological spin-offs using hierarchical information in international patent classification," *Technovation*, vol. 100, p. 102192, Feb. 2021.
- [2] Z. Liu, H. Kan, T. Zhang, and Y. Li, "DUKMSVM: A Framework of Deep Uniform Kernel Mapping Support Vector Machine for Short Text Classification," *Appl. Sci.-Basel*, vol. 10, no. 7, p. 2348, Apr. 2020
- [3] F. Sun and S. Lian, "nu-Improved nonparallel support vector machine," *Sci Rep*, vol. 12, no. 1, p. 17855, Oct. 2022.
- [4] S. Xu, "Bayesian Naïve Bayes classifiers to text classification," *Journal of Information Science*, vol. 44, no. 1, pp. 48–59, Feb. 2018.
- [5] H. Gao, X. Zeng, and C. Yao, "Application of improved distributed naïve Bayesian algorithms in text classification," *J. Supercomput.*, vol. 75, no. 9, pp. 5831–5847, Sep. 2019.
- [6] S. Jiang, G. Pang, M. Wu, and L. Kuang, "An improved K-nearest-neighbor algorithm for text categorization," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1503–1509, Jan. 2012.
- [7] F. Barigou, "Improving K-Nearest Neighbor Efficiency for Text Categorization," *Neural Netw. World*, vol. 26, no. 1, pp. 45–65, 2016.
- [8] U. A. K. Santoshi, "Text Classification Using Machine Learning Techniques," *Specialusis Ugdymas*, vol. 1, no. 43, Art. no. 43, Sep. 2022.
- [9] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
- [10] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, Austin, Texas, Jan. 2015, pp. 2267–2273.
- [11] P. Liu, X. Qiu, and X. Huang, "Recurrent Neural Network for Text Classification with Multi-Task Learning," May 2016.
- [12] L. Yao, C. Mao, and Y. Luo, "Graph Convolutional Networks for Text Classification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, Art. no. 01, Jul. 2019.
- [13] J. Li, G. Huang, J. Chen, and Y. Wang, "Dual CNN for Relation Extraction with Knowledge-Based Attention and Word Embeddings," *Computational Intelligence and Neuroscience*, vol. 2019, p. e6789520, Jul. 2019.
- [14] S. O. Alhumoud and A. A. Al Wazrah, "Arabic sentiment analysis using recurrent neural networks: a review," *Artif Intell Rev*, vol. 55, no. 1, pp. 707–748, Jan. 2022.
- [15] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," arXiv, May 19, 2014.
- [16] X. She and D. Zhang, "Text Classification Based on Hybrid CNN-LSTM Hybrid Model," in *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, Dec. 2018, vol. 02, pp. 185–189.
- [17] C. Yang, Y. Guo, X. Li, and B. Chen, "A Novel Method Using Local Feature to Enhance GCN for Text Classification," in *2021 11th International Conference on Intelligent Control and Information Processing (ICICIP)*, Dec. 2021, pp. 59–65.
- [18] S. Silvestri, F. Gargiulo, M. Ciampi, and G. De Pietro, "Exploit Multilingual Language Model at Scale for ICD-10 Clinical Text Classification," in *2020 IEEE Symposium on Computers and Communications (ISCC)*, Jul. 2020, pp. 1–7.
- [19] F. Chen, Z. Yuan, and Y. Huang, "Multi-source data fusion for aspect-level sentiment classification," *Knowledge-Based Systems*, vol. 187, p. 104831, Jan. 2020.
- [20] Z. Lu, P. Du, and J.-Y. Nie, "VGCN-BERT: Augmenting BERT with Graph Embedding for Text Classification," in *Advances in Information Retrieval*, Cham, 2020, pp. 369–382.
- [21] S. Zheng and M. Yang, "A New Method of Improving BERT for Text Classification," in *Intelligence Science and Big Data Engineering. Big Data and Machine Learning*, Cham, 2019, pp. 442–452.
- [22] W. E. N. Chaodong, Z. Cheng, R. E. N. Junwei, and Z. Yan, "Patent text classification based on ALBERT and bidirectional gated recurrent unit," *Journal of Computer Applications*, vol. 41, no. 2, p. 407, Feb. 2021.
- [23] A. H. Roudsari, J. Afshar, C. C. Lee, and W. Lee, "Multi-label Patent Classification using Attention-Aware Deep Learning Model," in *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Feb. 2020, pp. 558–559.
- [24] Y. Liu, P. Li, and X. Hu, "Combining context-relevant features with multi-stage attention network for short text classification," *Computer Speech & Language*, vol. 71, p. 101268, Jan. 2022.
- [25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv, May 24, 2019.

# IoT Technology for Intelligent Management of Energy, Equipment and Security in Smart House

Fangmin Yuan<sup>1</sup>, Yan Zhang<sup>2</sup>, Junchao Zhang<sup>2</sup>

School of Art and Design, Zhengzhou University of Industrial Technology, Zhengzhou, Henan, 451100, China<sup>1</sup>

School of Mechanical and Electrical Engineering, Zhengzhou University of Industry Technology, Zhengzhou Henan, 451100, China<sup>2,3</sup>

**Abstract**—The Internet of Things means that many of the daily devices used by humans will share their functions and information with each other or with humans by connecting to the Internet. The most important factor of the Internet of Things is the integration of several technologies and communication solutions. Identification and tracking technologies, wired and wireless sensors and active networks, protocols for increasing communication and intelligence of objects are the most important parts of the Internet of Things. In this article, an attempt has been made to determine the parts that can be used to make a house smart among the concepts and technologies related to web-based programs based on Internet of Things technology. Since it is very time-consuming to investigate the effect of all the Internet of Things technologies in smart homes, by studying and examining various types of research, the web-based program based on the Internet of Things is selected as an independent variable, and its effect on smart home management is investigated. For this purpose, a web-based program based on the Internet of Things for intelligent building energy management, intelligent equipment management, and intelligent security has been designed and implemented. As experimental results shown the proposed method the proposed method achieves better results compared to other existing methods in energy consumption by 33.8% reducing energy usage.

**Keywords**—Internet of things technology; smart homes; intelligent energy management; fuzzy logic

## I. INTRODUCTION

At the end of the 20th century, with the development of smart technologies, the development of communication networks and the Internet, the development of sensor networks and sensors, extensive efforts and studies began to use these technologies in order to provide solutions to improve human life [1-5]. One of the important applications of these technologies was communication with objects and obtaining information through these objects. This paradigm was presented for the first time by Kelvin Ashton in 1998 in a mockery. In fact, solutions are presented that could be used to communicate with anything through the Internet at anytime and anywhere and to identify them in the network. Access to environmental information and its status provided new forms of communication between people and objects and even between the objects themselves. It led to the introduction of the Internet of Things, which, in addition to people and information, also contained objects [6,7]. The definition of objects, according to European research projects on the Internet of Things, are all active participants in business, information,

and processes that can interact and communicate with each other and with the surrounding environment and exchange data and information in the environment. They deal with sensitive things, and they also have the possibility of reacting to real and physical world events. Objects have an effective role in the running processes. They also have the possibility of creating actions and services with or without direct human intervention. The Internet of Things is one of the new technologies that can be used in exploration and production processes, refineries, petrochemicals, pipelines, transportation, and distribution. This technology in the industry increases employee security, identifies health and safety issues, optimizes production, tolerates errors, and reduces operating costs. After the publication of the Internet of Things as a solution to communicate with objects and collect information from them, architectures were presented to establish and implement this solution.

IoT, which also includes machine-to-machine (M2M) communication, is a new technology that is used to connect all objects through the Internet for remote measurement and control. The Internet of Things uses several other technologies, such as wireless sensor networks, robotics, Internet technology, and smart devices. Conceptually, Internet of Things technology means a network of real-world entities, each of these entities or objects has a unique address and communicates and interacts with each other based on standard communication protocols. In fact, the Internet of Things technology makes it possible for objects around humans or objects in a building and house to exchange information with each other virtually, and by creating synergy; it causes a significant growth in optimal energy consumption and human life quality [8-10]. Currently, Internet of Things technologies are available, and its services and achievements can be used in various fields. Knowing the capacity and potential of the above technology is very important. While choosing the correct path to use in business and life, its opportunities and threats should be considered. Intelligence is one of the topics that has attracted the attention of many researchers in all branches of the modern science in today's world. Even today, cities can have the concept of intelligence. A smart city is where traditional networks and services are made flexible, efficient, and sustainable using information, digital, and communication technologies to improve city operations for the benefit of its residents and create a greener, safer smart city. They are more, faster, and friendlier [11-14]. These components make cities smart and efficient. Information and communication technology (ICT) is the key to transforming traditional cities into smart ones. Two

technological infrastructures named IoT and big data (BD) make smart cities efficient and responsive.

The construction industry is not exempted from this category. Significant actions have come to the fore during the last three decades. One of the most important developments in the construction industry is the issue of smart homes, which is manifested in residential and non-residential buildings of the 21st century with the advancement and integration of new technologies and with the help of electronic, computer, and network sciences. The possibility of its increasing development during the coming years has been provided. A collection of tools, techniques, and technologies for better management and control of homes to increase the comfort and peace of the home space is referred to as a smart home. In general, a smart home is a type of building equipment to create a pleasant environment that can turn the home space into a building with higher security, optimal energy consumption, and more security [15-18]. Nowadays, the development of smart homes to help people live more comfortably has been the focus of many researchers. Helping elderly and disabled people, creating more security [19], energy consumption management [20], and medical care [21] are among the motivations for the development of such environments.

The main work and the items needed for the temporary implementation of this research work is a browser that can communicate directly with an active device on the Internet. Devices that can cooperate have the ability to communicate with sensors, actuators, and other devices. One of the important features of web-based programs is their usability. The web pages in these programs are not simple pages that only contain a series of colors and simple images; each of these pages works like a web-based program, and due to the quick modifications, that can be made in the usability of the content of these pages. The web has happened. The content of web-based programs can be presented to users more dynamically than before. These programs extract data online from various sources and display it on web pages through browsers. Also, the growth and expansion of the smart grid are accompanied by fundamental changes in the distribution networks. These changes cover different areas from consumers to distribution companies and electricity retailers. On the one hand, electric energy retailers confront consumers with real-time or near-real-time changes in the price of delivered electric energy. On the other hand, with the growth and advancement of technology, scattered products in home sizes and with an acceptable initial cost will be available to consumers. Therefore, home consumers face a problem in the time and amount of use of the electric energy sources they need. In addition to this issue, the time of using the required household appliances during the day, taking into account the level of consumer satisfaction, adds to the complexity of the problem. With the growth and expansion of factors in smart electric networks, this article considers the design of a decision factor to solve this issue. The purpose of this factor is the optimal use of various sources of electric power supply available to the consumer, including sources of distributed household production and electricity purchased

from the grid, taking into account the level of satisfaction of the consumer from the time of using their household appliances [22-25].

In control systems, different approaches are used to receive information, measure the situation, and how make decisions and react. One of the widely used and common approaches in control systems is the fuzzy approach and fuzzy logic. Fuzzy logic is a form of logic used in expert systems and artificial intelligence applications [26,27]. Fuzzy logic is more general than other conventional logical approaches. The main factor of this generality can be seen in having the attitude of solving a complex problem in the form of a search space and making a decision based on the status of the question, answer, and control. This logic provided a basis for developing new tools, interacting with natural language, and displaying knowledge. An automatic lighting control system automatically turns off unnecessary lighting sources. Automatic lighting control systems by turning on and off lighting sources at different times and situations, as well as reducing the brightness of the lighting source based on the state of external factors around that source, will reduce energy consumption and increase the lifespan of lighting equipment [28-30]. Conventional control design methods use mathematical models to develop control systems and controllers. Fuzzy control refers to the concept of control from the point of view of linguistic description. Standard fuzzy logic can be easily applied and used in industry and industrial applications because it has a simple control structure, and its design is not a complicated and costly process. Based on fuzzy logic studies, dynamic fuzzy logic control has a more suitable performance than fuzzy logic control. In order to control the controller, several dynamic fuzzy technologies have been proposed, as examples of which we can point out the regulation of membership, the regulation of the input/output scale factor, and the regulation of descriptive rules. In this study, a fuzzy control system will be used to regulate the input and output of electricity and control the voltage of the smart home. For this purpose, after examining the fluctuations in the input voltage of the house, the designed system decides to turn off or light the house in such a way that the goal is to create maximum stability of the activity of the smart home components. The main goal is to design and implement a smart building management system using Internet of Things technology.

The main objectives of this study are as follows:

- 1) To develop a method to determine the parts that can be used to make a house smart among the concepts and technologies related to web-based programs based on Internet of Things technology.
- 2) To implement a sensor-based automation system based on the Internet of Things for intelligent building energy management, intelligent equipment management, and intelligent security has been designed and implemented.
- 3) To use a fuzzy inference system to control smart home operations in Internet of Things technology.

## II. LITERATURE REVIEW

The improvement of energy management, home security, environmental control, and other areas of smart home automation systems has been the focus of several research publications and existing literature. For analysis, prediction, and classification purposes, machine learning techniques have also been used in the IoT sector. The articles in this section of the IoT library address machine learning applications in intelligent systems and smart home automation.

A smart home automation system leveraging IoT technologies was presented by Govindraj et al. [34] to replace the current home automation system. Through the use of a satellite station and a radio frequency transceiver, the suggested system employs an Android application to monitor and regulate household appliances, temperature, motion, and gases. The ThingSpeak cloud platform is used to store sensor data. The essential commands for home control are delivered by a base station. Additionally, a mobile application with a graphical representation of sensor readings was developed to connect to the satellite station, base station, and cloud server for overall control of the house.

Rani et al. [35] suggested an artificial intelligence (AI) and natural language processing (NLP)-based voice-controlled home automation system. Voice commands are sent to a mobile phone to operate household appliances, and the phone's built-in natural language processing system decodes them. The system was simply utilized to operate household appliances; it was not expanded to include other parts of home automation like control, monitoring, and motion and intrusion detection.

A low-cost smart home system design and prototype implementation was presented by Aadel and Ali [36]. The system was made to regulate the temperature, power, lights, and doors in the house. An Arduino board, servo motors, LED lights, temperature, and motion sensors were used to demonstrate a prototype implementation of the system. The INA219 high-side DC sensor was used to monitor the drop and supply of power, the DHT11 temperature and humidity sensor was used to measure the temperature and humidity in the house, and the door and windows were controlled by servo motors. The communication method employed by the system was Bluetooth, which has a limited range of communication.

Parsa et al. [37] suggested a system for the optimal and autonomous regulation of electrical home equipment. By automatically turning on and off the smart plugs connected to certain home appliances at the proper times, the suggested system is designed to reduce power usage. To choose the ideal period of use prior to implementing the automatic switch in accordance with the established criteria, an optimization approach was applied. The system's design gives the supplier the upper hand over the household's electrical user.

The design of an embedded smart house control and monitoring system employing an STM32 microcontroller was presented by Xiaodong and Jie [38]. Their technology was created for home control and interior environmental conditions monitoring (temperature, humidity). To communicate remote control and Zigbee terminal connectivity between the home gadgets, the authors used a GSM and GPRS module. Their

method was implemented using a combination of coordinated remote control, feedback, and embedded real-time operating system (COS-II) for home appliances.

## III. PROPOSED METHOD

The proposed load model assumes that each user has different devices with different energy needs, power demands, and working hours. Equipment can be divided into two categories: permanent load and temporary load, and each load model have its characteristics. Devices with constant load (refrigerator, heating, cooling, water heater, electric kettle, air conditioning, etc.), this type of load are known as energy consumption/production, which covers the entire time interval of the energy simulation program. Temporary load devices, this type of load is known by the start time and end time of the operation. This subcategory can be divided into two other subcategories such as (TV, lighting, cooking, hair dryer, etc.). The first one represents the devices that are the main choice of users and cannot be used backward (such as laptop chargers, washing machines, etc.), which is a "transferable burden". The second subcategory is more flexible and can be used at other times. Each of these two load models can be control or binary devices. Control devices are devices with different states to stay on, and binary devices have only two states, on or off. The proposed system of this research combines different factors that react to their environment based on a set of predetermined instructions. An agent is an entity that is created to perform tasks. An intelligent agent is an autonomous entity that perceives its environment through sensors and responds to the environment using stimuli [31]. This agent tries to maximize his desired productivity.

### A. Coordinating Agent Modeling

Considering a set of  $A=\{a_1, a_2, \dots, a_N\}$  devices in which each device has its proportional energy consumption. For each  $a \in A$ , the energy consumption vector of  $Y_a$  is calculated as equation (1).

$$Y_a = [y_a^1 \dots y_a^H] \quad (1)$$

Where  $H=24$  hours,  $H$  is the timing range that indicates the number of hours per day that can be used by  $a \in A$  device. The coordinating agent considers the computational domain  $H$  to decide on the energy consumption schedule.  $y_a^j (j \in \{1..H\})$  represents the energy consumption for a device ( $a$ ) in one hour. For each  $a \in A$ , the user specifies the convenience of the  $(a_a < \beta_a)$   $a_a, \beta_a \in H$  range, while the beginning and end of a time make it possible to schedule energy consumption for the device. But  $(\beta_a - a_a)$  must be greater than the time required for  $(t_a^{req})$  to complete the normal process of the device ( $a$ ) as described in equation (2).

$$\beta_a - a_a \geq t_a^{req} \quad (2)$$

where  $a$  is the start time of the device operation,  $\beta$  is the end time of the device operation. Also, the total daily energy consumption for each device is defined as equation (3).

$$EC_a^o = \sum_{h=a_a}^{\beta_a} y_a^h \quad (3)$$

$y_a=0$  is expected to be for all  $h>\beta$  and  $h<a_a$ . Upper and lower limits have been specified for  $y_a$ , which are used in the selection of the energy consumption scheduling vector.

$$\delta_a^{\min} \leq y_a \leq \delta_a^{\max} \quad (4)$$

where  $\delta_a^{\min}$  is the minimum standby power level, and  $\delta_a^{\max}$  is the maximum power level. For each center, the total energy consumption per hour must be less than or equal to the predetermined energy threshold ( $E^{\max}$ ), as calculated in eq. (5).

$$EC^{oTotal} = \sum_{a \in A} y_a^h \leq E^{\max}, \forall h \in H \quad (5)$$

where  $E^{\max}$  and  $EC^{oTotal}$  are used as the input/output of the fuzzy system, respectively as an index of the fuzzy threshold limit and power demand. By combining eq. (3) to (5), energy consumption timing vector choices are determined. Therefore, the timing set  $Y$  for all  $y_a$  can be determined as eq. (6).

$$\left\{ \begin{array}{l} y | EC_a^o = \sum_{h=a_a}^{\beta_a} y_a^h, \\ \forall a \in A, \beta_a - a_a \geq t_a^{req}, \\ \delta_a^{\min} \leq y_a \leq \delta_a^{\max}, \\ \forall a \in A, h \in [a_a, \beta_a], \sum_{a \in A} y \leq E^{\max} \end{array} \right. \quad (6)$$

where  $y=(y_a; A_a \in A)$  represents the energy consumption scheduling vector that contains all the variables for all devices. Therefore, a vector  $y$  is true only when  $y \in Y$  is true.

### B. Permanent Factor Modeling

The permanent factor is related to any permanent load device. In this subcategory, loads flow regularly and depend on the device's internal temperature. Also, the comfort range depends on the high and low levels of  $[T_{ac}^{\min}, T_{ac}^{\max}]$  temperature. The permanent agent tries to maximize the satisfaction function of each permanent load device.

$$SF = \left\{ T \left| \begin{array}{l} SF(T) = SF(T_{ac}^{\min}) = SF(T_{ac}^{\max}) = 100, \\ \forall T \in [T_{ac}^{\min}, T_{ac}^{\max}] \end{array} \right. \right\} \quad (7)$$

A constant load device's satisfaction function depends on its descriptive variable. For example, the air conditioning service is dependent on its temperature (T), which is seen in Fig. 1, and a user is satisfied if the temperature of the room in which he is sitting is 21°C to 22°C.

In order to avoid the peak load demand without affecting the user's comfort, the permanent agent uses the scheduling operation method, shown in relations (4) and (5). The flexibility of this service comes from the possibility of editing the quantities of energy consumed/produced in all periods. Therefore, it increases or decreases the factor with the assumed limits  $[\delta_a^{\min}, \delta_a^{\max}]$  of each device.

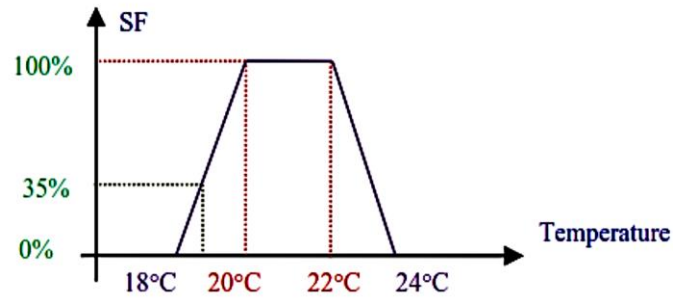


Fig. 1. Depending on the satisfaction of air ventilation

### C. Temporal Agent Modeling

The temporary factor is related to any temporary load device. Mandatory execution load starts when the user requests execution. Since the force consumption is constant, there is no other option but to regulate its operations in the normal force demand and these tasks do not need scheduling. However, in periods of high-power demand of the provider, the operator is instructed to use the fuzzy system to control the required load operations. The transferable load starts its duty when the temporary factor (such as the dishwasher) is according to the comfort zone of the user  $[a_a, \beta_a]$  and the limit of the relationship (3) of the operation of each device. The satisfaction function of a temporary load device is dependent on the transfer time in service  $[a_a, \beta_a]$ , which is also dependent on the optimal time of the user's request, shown in equation (8). For example, the user wants his clothes to be cleaned at 9:30 AM, which can be seen in Fig. 2.

$$SF = \left\{ t \left| \begin{array}{l} SF(t) = SF(a_a) = SF(\beta_a) = 100, \\ \forall t \in [a_a, \beta_a] \end{array} \right. \right\} \quad (8)$$

The temporary agent used the existing list of priorities prepared by the coordinator and then used three types of fuzzy logic based on the control strategy as an efficient solution to transfer the required power of the devices in periods of high demand.

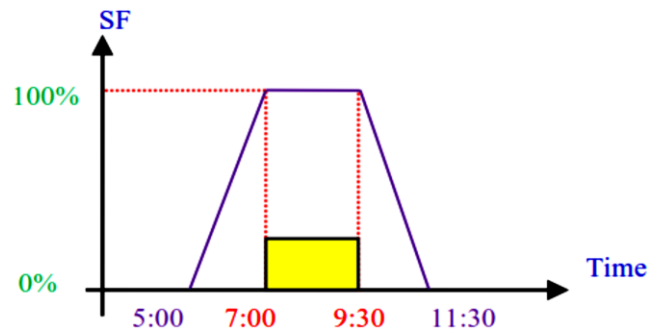


Fig. 2. Washing machine satisfaction function

## IV. PROPOSED FUZZY INFERENCE SYSTEM

In this section, the proposed fuzzy inference system is discussed. Principally the controlling of smart devices in IoT systems is not contained absolute value to control the various processes. Therefore many-value with approximate number ids

required to use in these systems Fuzzy logic is a form of many-valued logic in which the truth value of variables may be any real number between 0 and 1. It is employed to handle the concept of partial truth, where the truth value may range between completely true and completely false [33].

Fig. 3 shows the block diagram of the proposed fuzzy inference system.

The proposed method has two inputs and one output variable. In the fuzzification stage, the fuzzy controller receives the inputs and maps them to their membership functions, called fuzzy sets. The degree of membership in the fuzzy set  $\mu$  for the input  $x$  is determined in the fuzzification step. This degree of numerical membership is between 0 and 1. The value 0 means that  $x$  is not a member of the fuzzy set; the value 1 means that  $x$  is completely a member of the fuzzy set. The values between 0 and 1 specify the fuzzy members, each of which partially

belongs to the fuzzy set. The membership function (MF) is a curve that specifies how to map each point in the input space to the membership value (or degree of membership) between 0 and 1. Sometimes the entrance space of the community is also called the entrance. The most used shapes for membership functions are triangular, trapezoidal, and Gaussian. In the proposed method, membership functions assigned to input and output variables are considered triangular. The edges of the triangle can be specified by the triplet  $(a, b, c)$  (where  $a < b < c$ ). The parameters  $\{a, b, c\}$  of the  $x$  coordinates of the three edges of the desired triangular function are specified. Fig. 4 shows a triangular membership function  $(2, 4, 6)$ . Point 4 has the largest value in the membership function.

In this study, the fuzzy set includes a maximum of 4 states, which are different for each input membership function, which is explained in the following sections.

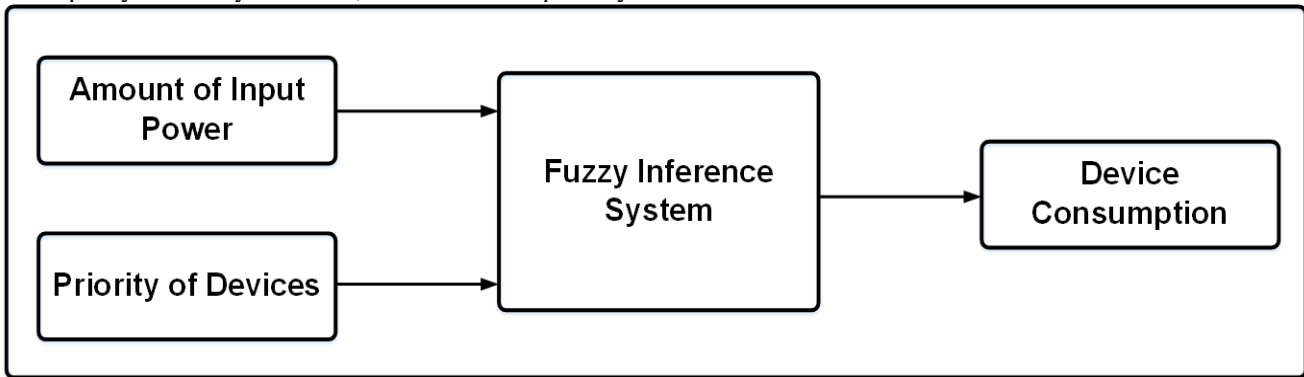


Fig. 3. Proposed fuzzy inference system.

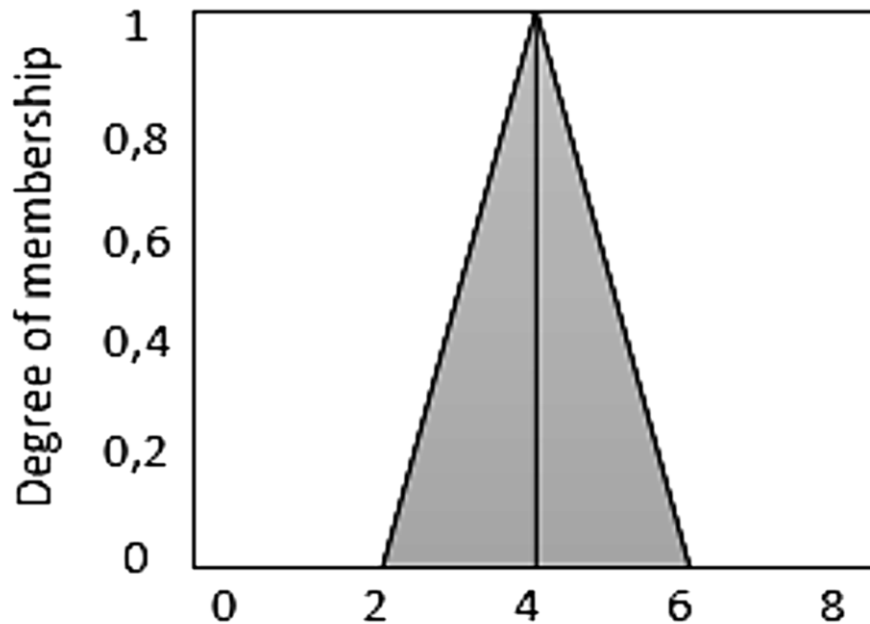


Fig. 4. An example of a triangular membership function.

#### A. Input Power Membership Function

It is assumed that the number of input power modes equals (normal, less than normal, medium, and weak). The value of 0

indicates that the input power is weak, and more care should be taken regarding its distribution. While the value of 500 indicates the normality of the input power. The triangular membership function maps the number of input power states

from 0 to 500 watts according to the degree of membership into 4 fuzzy sets (Weak, Medium, Less Than Normal, and Normal). This allocation is shown in Fig. 5.

According to Fig. 5, the fuzzy sets of this function are {Weak: triangle [0 0 1800]}, {Medium: triangle [0 1800 3500]}, {Less Than Normal: triangle [1800 3500 5000]}, {Normal: triangle [3500 5000 5000]}.

### B. Priority Membership Function

The Priority membership function determines the priority of the device, which in this research includes (low, medium, important, and very important) states. A value of zero for this function indicates a low priority state, a value of 1 indicates a medium priority, a value of 2 indicates an important priority, and 3 indicate a very important state. The triangular membership function maps the number of device priority states from 0 to 3 according to the degree of membership into 4 fuzzy sets (Low, Medium, Important, Very Important).

This assignment is shown in Fig. 6. The fuzzy sets of this function are {Low: triangle [0 0 1]}, {Medium: triangle [0 1 2]}, {Important: triangle [1 2 3]}, {Very Immortan: triangle [2 3 3]}.

### C. Device Usage Membership Function

The Device Usage member function specifies the amount of device usage. In this study, the Device Input Type variable is determined, which includes (control and binary) modes. Suppose it is a control device such as a fan. In that case, the amount of consumption of the device is decided by reducing the input power and the priority of the device. This type's device consumption includes (off, low, medium, and high) modes. For example, a ceiling fan with a consumption value of [65 75], a value of zero for this function indicates that it is off, a value of 65 low consumption and 120 consumption It is average, and the value of 175 shows a high state. The triangular membership function maps the number of device usage states to 4 fuzzy sets (Off, Low, Medium, High) according to membership degree. This allocation is shown in Fig. 7. The fuzzy set of this function is {Off: triangle [0 0 65]}, {Low: triangle [0 65 120]}, {Medium: triangle [65 120 175]}, {High: triangle [120 175 175]}.

If the type of input device is binary, we will have only two states, off and on, which determine the amount of consumption of the device. In this case, the Device Usage membership function will be Fig. 8.

The fuzzy set of this function in this type of device is {Off: triangle [0 0 1]}, {On: triangle [0 1 1]}. The input variable states with these diagrams do not change quickly from one state to the next. Instead, when the input changes, it loses a value in one membership function while gaining another value in the next state. In other words, an input variable with a specific membership degree is a part of two membership functions. For example, in Fig. 7, when the Input Power is equal to 200 watts, the input completely belongs to the membership function of Medium. However, when the Input Power is equal to 280 watts, the input is partially (0.5 each) a part of two membership functions, Medium and Important. Generally, a fuzzy system is formed based on human

experience and expert information. The limits of the modes are also defined in the same way.

### D. Fuzzy Inference System

An inference engine is equipped with fuzzy rules to decide on an output channel based on the current state of the network. An inference engine with a set of linguistic statements to describe the system using the number of conditional rules (if-then) where the if-then part is specified by (precedence) and the then part is specified by (tail). Usually, the knowledge of an expert is used to form the rules of the fuzzy inference system. Table I contains rules used in the proposed method for control devices with two fuzzy inputs and one fuzzy output. This table contains different output values for different input ranges. Filling a data table with fuzzy attributes (scalability) is a matter of taste.

Table II contains rules that are used in the proposed method for binary devices with two phased inputs and one phased output. These tables are prepared based on basic knowledge about the effect of each criterion on the overall performance of the proposed method.

The fuzzy rules usually combine several assumptions using fuzzy operators such as fuzzy intersection (AND) and fuzzy union (OR). If the rule uses the AND relationship to map two input variables, at least these values are used as output. In contrast, for the OR relationship, the maximum is used. In the proposed method, the AND operator is used to combine the fuzzy inputs.

The example in Fig. 7 helps clarify the issue. Input Power and Priority parameters have values of 250 and 1.8, respectively. As shown in Fig. 9, Input Power is a part of the membership functions of Medium and Less Than Normal, and the part of each membership function is 0.5. The Priority entry is a part of Medium and Important membership functions, shown in Figure 10. In this case, the degree of membership for membership functions Medium and Important is 0.2 and 0.8, respectively. The fuzzy sets of this function are {Weak: triangle [0 0 0.33]}, {Medium: triangle [0 0.33 0.66]}, {Less Than Normal: triangle [0.33 0.66 1]} {Normal: triangle [0.66 1 1]}.

Fig. 9 and 10 shows the Input Power and Priority values, the four combinations between Input Power and Priority are shown in Fig. 11(a) Input Power: Medium and Priority: Important, Fig. 11(b) Input Power: Medium and Priority: Medium, Fig. 11(c) Input Power: Less Than Normal and Priority: Important, Fig. 11(d) Input Power: Less Than Normal and Priority: Medium.

### E. Composition and Defuzzification

Defuzzification is the process of producing a measurable result in fuzzy logic and converting the fuzzy control action to a definite value. The output of all rules must be collected and converted into a single output. Two methods are widely used for defuzzification. Method 1 Center of Gravity (CoG): This method finds the geometric center. Also, this method selects the output with the largest area. Method 2: Maximum mean (MoM): This method gives the values that have the maximum degree of membership according to the fuzzy membership function.



It is simpler but loses useful information, while CoG, which is used as a common method, is more efficient. CoG is our chosen defuzzification method to produce a definite value in this research. In the phase of defuzzification, the four output values obtained in Fig. 11 are combined, and a single output is extracted using the center of gravity method. As shown in Fig. 12, the fuzzy outputs of the membership function with the

same output are added together, while the values in different membership functions are taken together (that is, their maximum value is considered).

The output value can be calculated from eq. (9).

In this case, the output value is equal to eq. (10).

$$\text{Obtained DeviceUsage} = \frac{\text{Obtained DeviceUsage}}{\text{Degree of Membership Functions}} \tag{9}$$

$$\text{Obtained DeviceUsage} = \frac{(0.33 \times 0.2) + (0.33 \times 0.5) + (0.33 \times 0.2) + (0.66 \times 0.5)}{0.2 + 0.5 + 0.2 + 0.5} \cong 0.45 \tag{10}$$

$$EC^{oTotal} = \sum_{a \in A} y_a^{h=1} = 420 + 300 + 1800 + 1500 + 2000 = 6020W \tag{11}$$

$$ES(\%) = \left(1 - \frac{EUA_{After}}{EUB_{Before}}\right) \times 100 \tag{12}$$

$$EUB_{Before} = 420 + 300 + 1800 + 1500 = 4020W ,$$

$$EUA_{After} = 210 + 267.6 + 1274.2 + 666.7 = 2418.5W ,$$

$$ES = \left(1 - \frac{2418.5}{4020}\right) \times 100 = 39.8 \tag{13}$$

According to this formula, the degree of the membership function of each rule is multiplied by the maximum value in the membership functions in the output value and then divided

by the sum of all the values of the membership function. In this example, the power consumption of the device is set to Medium in Fig. 13.

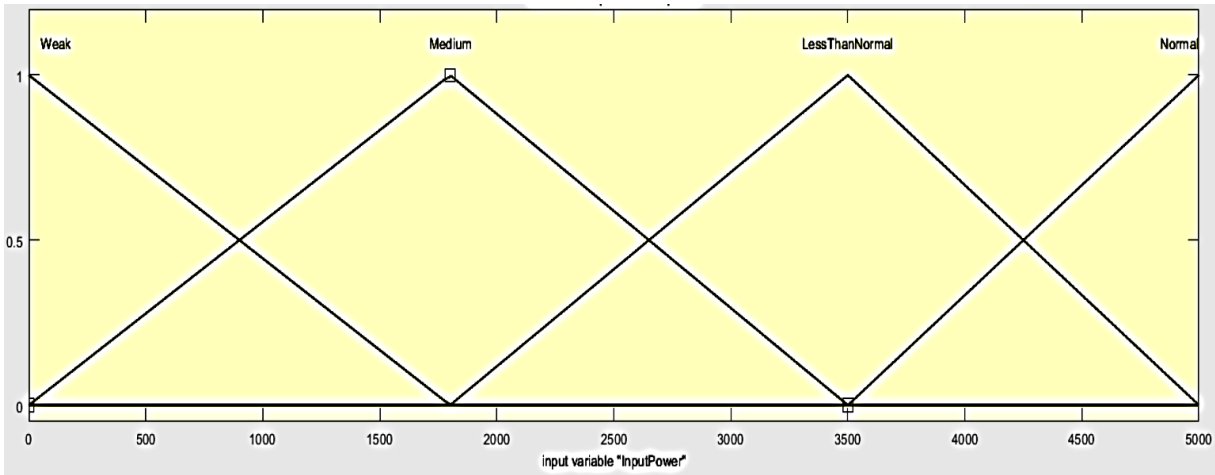


Fig. 5. Input power membership function

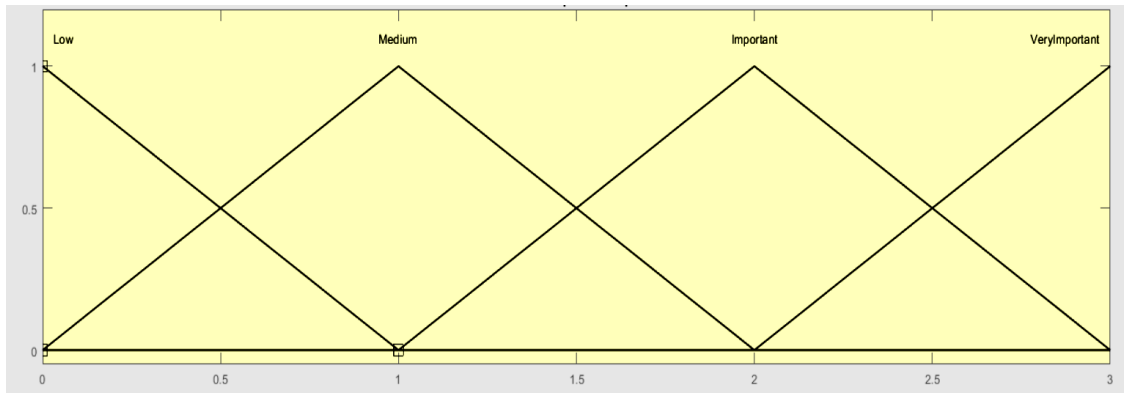


Fig. 6. Priority membership function

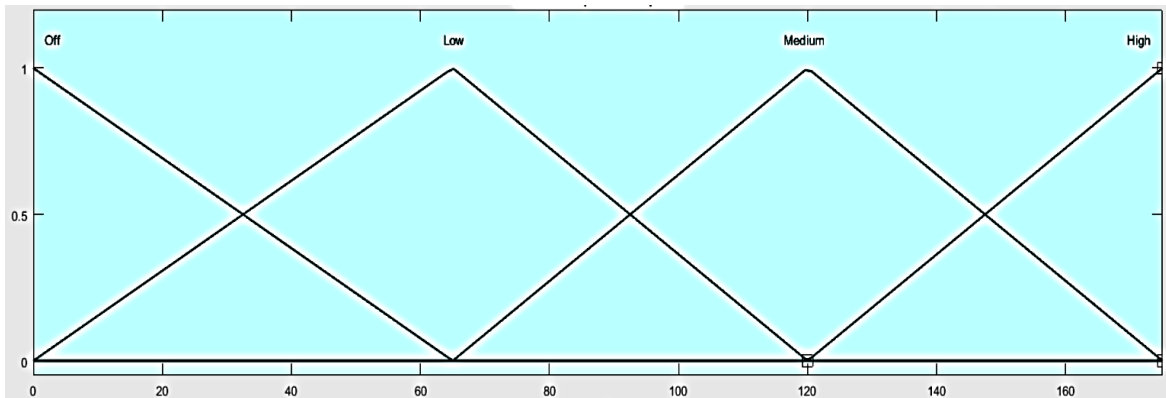


Fig. 7. Device usage membership functions in the control device type

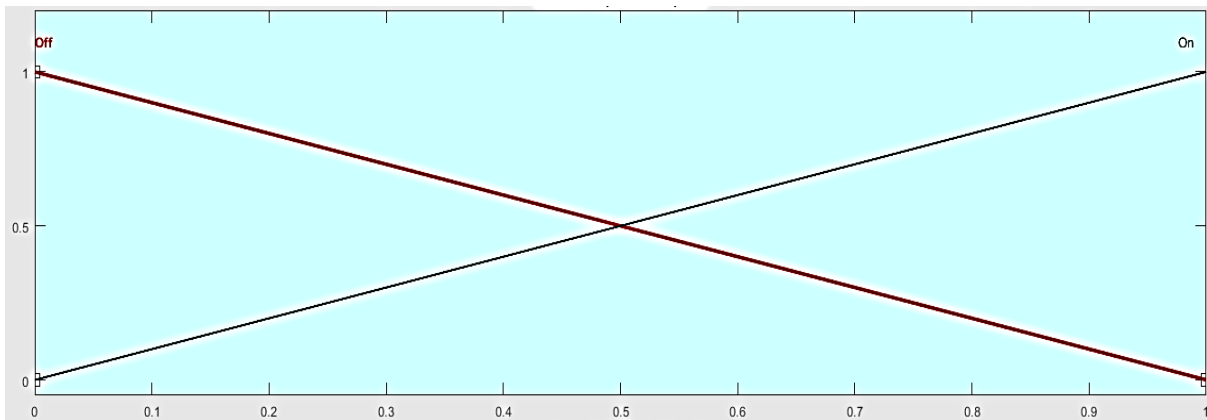


Fig. 8. Device usage membership functions in binary device type

TABLE I. INFERENCE RULES OF THE PROPOSED METHOD FOR CONTROL DEVICES

Results		Priority			
		Low	Medium	Important	Very Important
Input Power	Weak	Off	Off	Low	Low
	Medium	Off	Off	Low	Medium
	Less Than Normal	Low	Low	Medium	High
	Normal	High	High	High	High

TABLE II. INFERENCE RULES OF THE PROPOSED METHOD FOR BINARY DEVICES

Results		Priority			
		Low	Medium	Important	Very Important
Input Power	Weak	Off	Off	Off	On
	Medium	Off	Off	On	On
	Less Than Normal	Off	On	On	On
	Normal	On	On	On	On

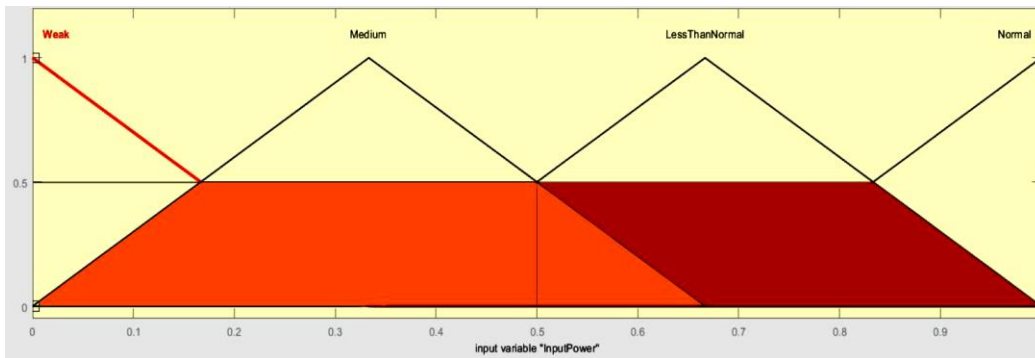


Fig. 9. Input power value is equal to 0.5

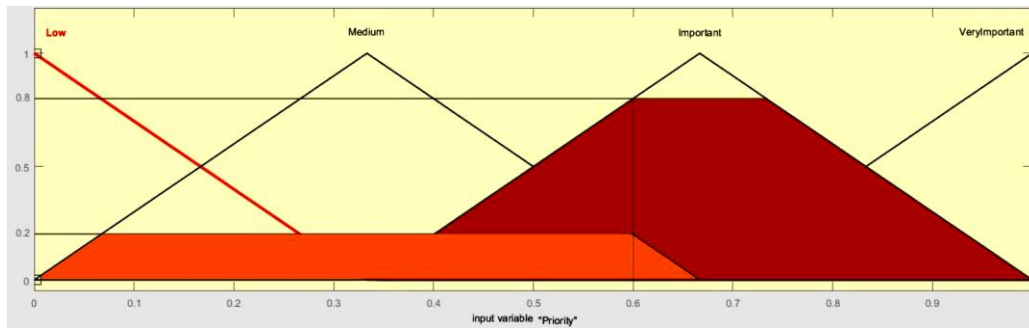


Fig. 10. Priority value is equal to 0.6

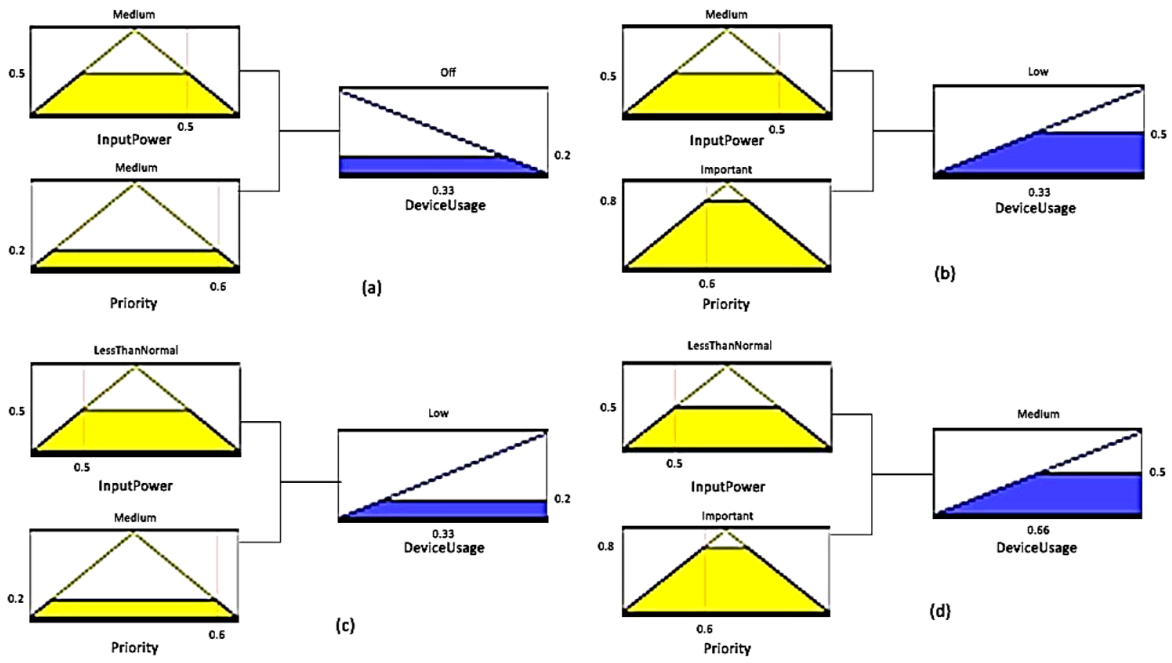


Fig. 11. Output for (a) rule1, (b) rule2, (c) rule3, (d) rule4, for control devices

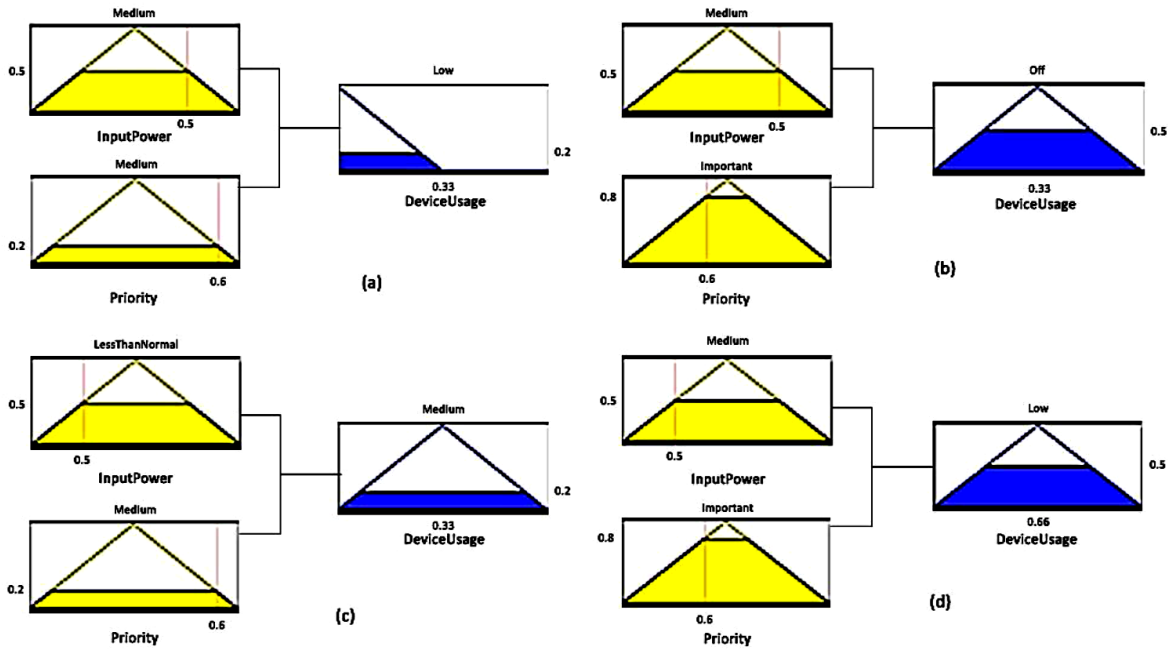


Fig. 12. Output for (a) rule1, (b) rule2, (c) rule3, (d) rule4, for binary devices

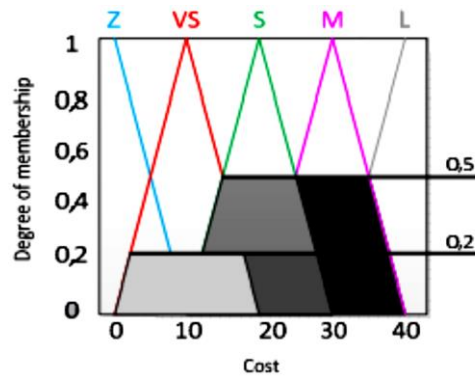


Fig. 13. Combining the membership function of device usage from all rules and generating a definite value using the CoG method.

### V. SIMULATION EVALUATION

In this section, the simulation results are evaluated. The simulation was done with MATLAB software. Each load group is coordinated with a scheduling policy to reduce energy by compressing power demand or postponing requests. In this section, some devices' power demand control behavior and the type of control used have been analyzed. Refrigerators and freezers can be used for a short time, provided the temperature is kept within a certain range. The agent can predict the energy needed for the next time by observing the T parameters and the time the door is open. The permanent agent uses the scheduling policy according to the comfort range of  $[T_{ref}^{min}, T_{ref}^{max}]$ .

The washing machine and dishwasher are binary devices with only two states, on or off. Therefore, the temporary agent predicts its consumption and gives users the price of electricity at the time of use, which the provider provides. The user can benefit from washing at night at a lower price. For example, the user expects the dishes to be ready by dinner, so he sets the time of use as  $\alpha=2$  PM and  $\beta=6$  PM. So, suppose it is at the peak of electricity consumption or during the electricity fluctuation. In that case, this device will be turned off and transferred to a more suitable time. Television is a device whose power level is flexible. The TV has three functions with three levels, so the temporary agent uses a scheduling policy. If the threshold is soon reached, and some devices cannot be stopped, the agent switches the TV's function to ambient light or standby mode. Be with the level. TV is a device whose power level is flexible. The TV has three functions with three levels, so if the threshold is reached soon, the temporary agent will use the timing policy. Some devices cannot be stopped, so that the agent will adjust the TV function to the ambient light or mode. Ready to work with level  $y_a = \delta_a^{min}$ . The water heater is very important when the user wants a quick shower. However, for the user who has planned his bath time, the agent can heat the water before the specified time. In addition, this factor can cut off the boiler's power for an hour without the user noticing. The permanent factor controls the daily energy of the water heater using the scheduling policy.

In this section, we analyze the power consumption control in three types of devices in the state of peak power demand in Fig. 14. The simulation was divided into 15-minute segments during the hours of the day. A set of  $(t \in Time)$  time segments was created in which  $Time = \{1, 2, 3, \dots\}$  etc.} is The minimum

and maximum power of the devices used in the strategy are refrigerator (140 to 420 W) control device, with priority 4, TV (100 to 300 W) control device, with priority 2, water heater (750 to 1800 W) control device, with Priority 3 and time satisfaction function 22-24, air heater (500 to 1500 watts) control device, with priority 3 and temperature satisfaction function 20-22 degrees Celsius, dishwasher (2000 watts) binary device, with priority 1 and time satisfaction function 14-18. Fig. 14 shows the output phase system of the refrigerator. Power below 140 watts will cause the refrigerator to turn off, and 300 watts will cause low consumption of the refrigerator, and the most consumed time of the refrigerator is 420 watts.

Fig. 15 shows the fuzzy system of TV output. Power below 100 watts will turn off the TV, and power of 200 watts will cause low consumption of the TV, and the most consumed time of the TV is equal to 300 watts.

Fig. 16 shows the output phase system of the water heater. Power below 750 watts will turn off the water heater, and 1400 watts will cause low consumption of the water heater, and the most consumed time of the water heater is 1800 watts.

Fig. 17 shows the output phase system of the air heater. Power below 500 watts will turn off the air heater, and 1000 watts will cause low consumption of the air heater and the most time-consuming air heater is 1500 watts.

Fig. 18 shows the fuzzy output system of the dishwasher. Power below 2000 watts causes the dishwasher to turn off, and power below 2000 watts causes the dishwasher to stay on.

In one hour, the house's energy is calculated based on equation (4) and should be a maximum of 3000 watts ( $E_{max}=3000$ ). If all the devices are on and have their maximum consumption, the total consumption power is calculated according to equation (11). The total power consumption is 6020 watts, which is more than the maximum consumption. Therefore, the coordinating agent predicts the peak power and performs some tasks in advance. The operation of the dishwasher can be moved to midnight. The TV will work in a low-power mode. The refrigerator can be used for one hour as long as the door is closed. The agent of the satisfaction function checks the water heater and the air heater, schedules their operations in case of peak power demand, and cuts off the devices with a low priority. The power consumption output by the proposed system is shown in Fig. 19.

In the first 15 minutes of the simulation, the refrigerator consumes 210 watts, the television consumes 267.6 watts, the water heater consumes 1274.2 watts, the air heater consumes 666.7 watts, and the dishwasher consumes zero watts. Therefore, the TV and dishwasher are turned off, and the refrigerator, water heater, and air heater are on with low consumption. In 15 minutes of the 20th simulation, the refrigerator consumes 250 watts, the television 150 watts, the water heater 900 watts, the air heater 1023.2 watts, and the dishwasher consumes 0 watts. Therefore, the dishwasher is turned off, and the refrigerator, TV, water heater, and air heater are on with low consumption. In the 56th 15 minutes of the simulation, which is the time dependent on the satisfaction of the dishwasher, the refrigerator consumes 140 watts, the television 100 watts, the water heater zero watts, and the air heater 750 watts, and the dishwasher consumes 2000 watts. Therefore, the water heater is turned off, the refrigerator and air heater are on with low consumption, and the dishwasher started working for half an hour. After the dishwasher is finished, the

water heater is turned on and continues its operation. The reduction of 15 minutes is due to the reduction in water heater consumption. In all cases, the red line in Fig. 19 shows the amount of energy consumption without the proposed research method. As can be seen, the consumption of the proposed method is always lower than the method of the article [31]. The amount of energy saving can be obtained by equation (12).

Where  $EUA_{After}$  indicates the energy consumption after the proposed algorithm and  $EUB_{Before}$  indicates the energy consumption before the proposed algorithm. In fact, building energy saving is defined as the difference ratio of energy use before and after the implementation of the proposed algorithm in the same period. Energy saving has been calculated in two intervals. In the interval [1 1], the first 15 minutes, the total energy consumption is calculated in eq. (13). In this case, the force has been reduced by 39%. In [32 39], the total amount of consumed energy was equal to 2990, and the proposed method reduced the energy by 29.1%.

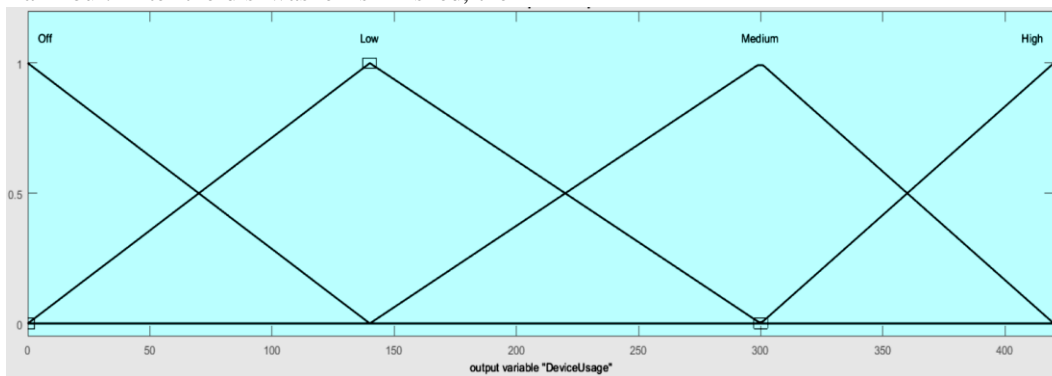


Fig. 14. Fuzzy system of refrigerator output

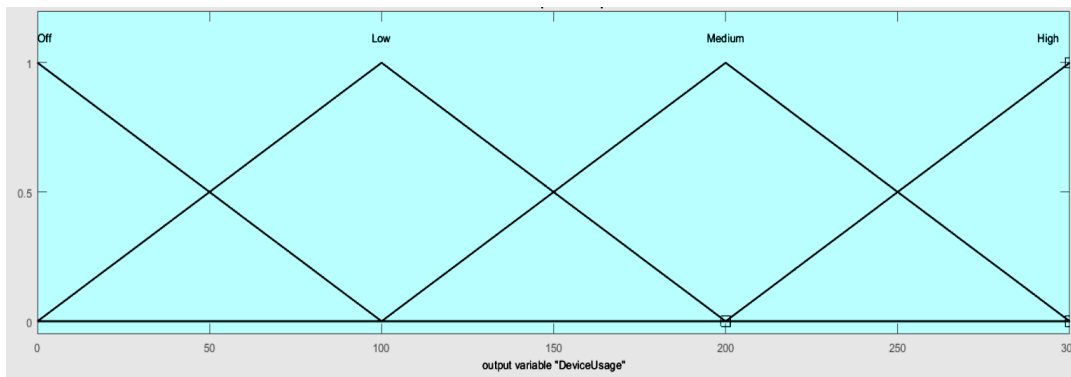


Fig. 15. Fuzzy TV output system

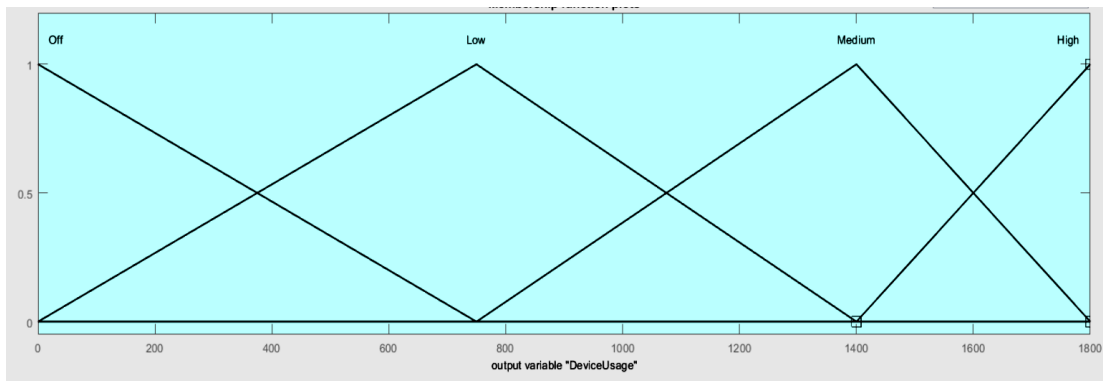


Fig. 16. Water heater output phase system

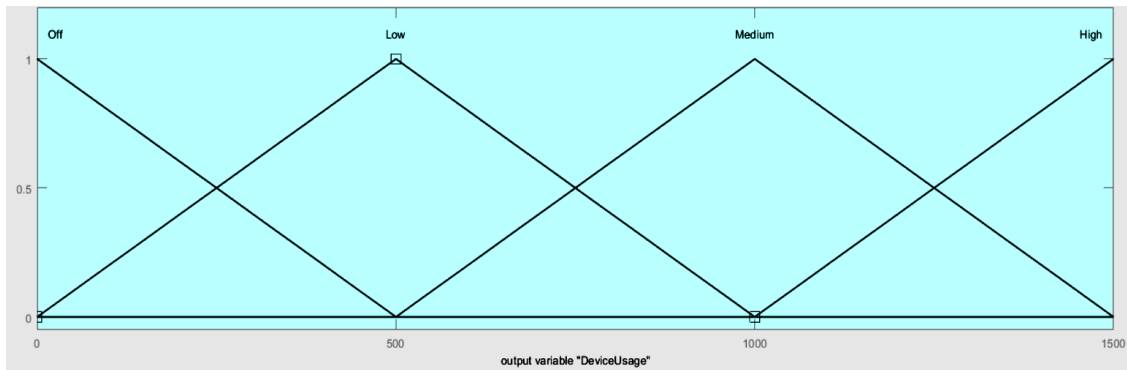


Fig. 17. Phased air heater output system

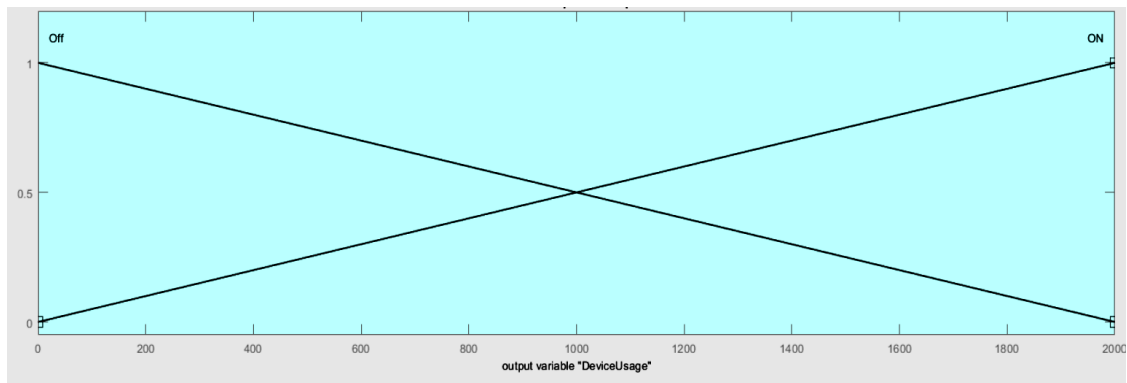


Fig. 18. Fuzzy output system of the dishwasher

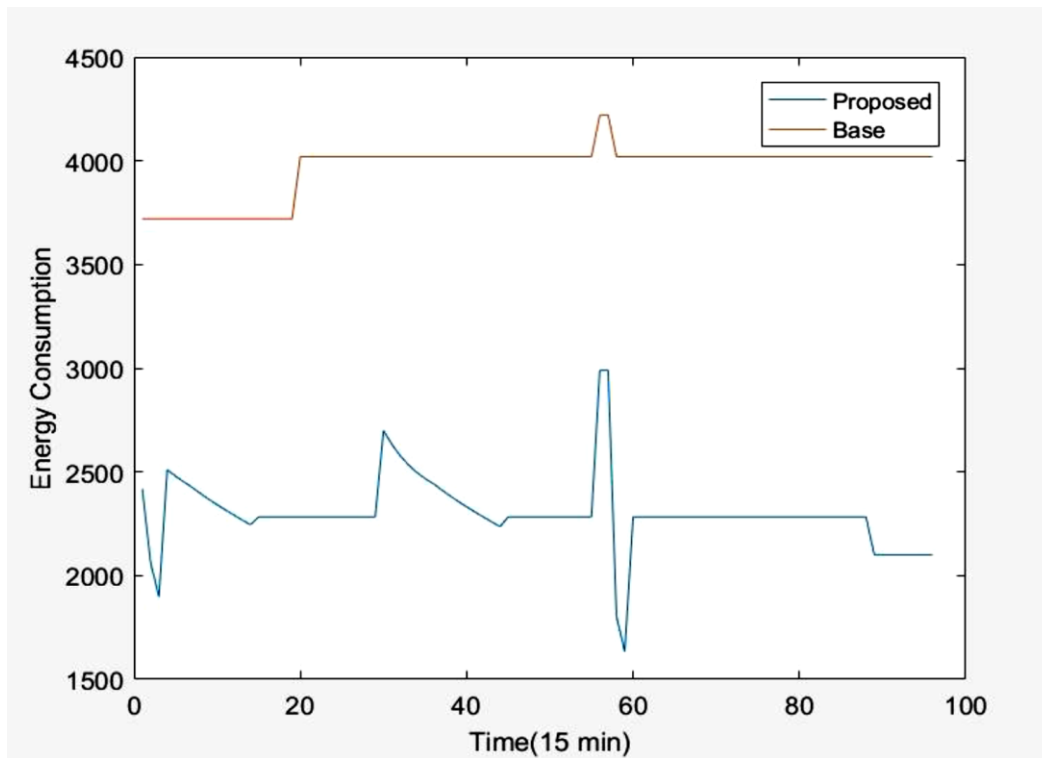


Fig. 19. Energy consumption based on 15 minutes out of 24 hours

## VI. CONCLUSION

The proposed load model assumes that each user has different devices with different energy needs, power demands, and working hours. Devices are divided into two subgroups: permanent load and temporary load, each of these two "load" electricity amount models can be controlled or binary devices. The proposed model of this research consists of two inputs named (amount of input power) and (device priority) and one output (device consumption). Decision-making for an output channel based on the current state of the network is equipped with fuzzy rules. The number of input power modes equals (normal, less than normal, medium, and weak). The priority of the device includes modes (low, medium, important, and very important). Suppose the device is of a control type. In that case, the device's consumption is decided by reducing the input power and the device's priority. This type's device consumption includes (off, low, medium, and high) modes. Suppose the type of input device is binary. In that case, we will have only two states, off and on, determining the device consumption amount. Building energy saving is defined as the difference ratio of energy use before and after implementing the proposed algorithm in the same period. This criterion has been used to evaluate the proposed method. According to comparison result among the proposed method with other existing methods, this method achieves better results compared to methods in energy consumption by 33.8% reducing energy usage.

**Conflicts of Interest:** The authors declare no competing interests.

**Data Availability Statement:** The authors do not have permission to share data.

## REFERENCES

- [1] A. Monacchi, D. Egarter, W. Elmenreich, Integrating households into the smart grid, 2013 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES), IEEE, 2013, pp. 1-6.
- [2] A. Jarrah Nezhad, T.K. Wijaya, M. Vasirani, K. Aberer, SmartD: Smart meter data analytics dashboard, Proceedings of the 5th international conference on Future energy systems, 2014, pp. 213-214.
- [3] V.T. Shi, D.R. Ngh, Channel Estimation Optimization Model in Internet of Things based on MIMO/OFDM with Deep Extended Kalman Filter, Advances in Engineering and Intelligence Systems, 1 (2022).
- [4] H. Ziekow, C. Goebel, J. Strüker, H.-A. Jacobsen, The potential of smart home sensors in forecasting household electricity demand, 2013 IEEE international conference on smart grid communications (SmartGridComm), IEEE, 2013, pp. 229-234.
- [5] S. Meiling, T. Steinbach, M. Duge, T.C. Schmidt, Consumer-oriented integration of smart homes and smart grids: A case for multicast-enabled Home Gateways?, 2013 IEEE Third International Conference on Consumer Electronics, Berlin (ICCE-Berlin), IEEE, 2013, pp. 279-283.
- [6] P. Finn, C. Fitzpatrick, Demand side management of industrial electricity consumption: Promoting the use of renewable energy through real-time pricing, Applied Energy, 113 (2014) 11-21.
- [7] A.J. Roscoe, G. Ault, Supporting high penetrations of renewable generation via implementation of real-time electricity pricing and demand response, IET Renewable Power Generation, 4 (2010) 369-382.
- [8] M. Aiello, G.A. Pagani, The smart grid's data generating potentials, 2014 Federated Conference on Computer Science and Information Systems, IEEE, 2014, pp. 9-16.
- [9] E. Colby, M. England, How will a Smart Grid manage Consumer energy appliances?, 2011 IEEE International Conference on Consumer Electronics (ICCE), IEEE, 2011, pp. 513-514.
- [10] S. Karnouskos, Crowdsourcing information via mobile devices as a migration enabler towards the SmartGrid, 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), IEEE, 2011, pp. 67-72.



- [11] I. Vorushylo, P. Keatley, N. Shah, R. Green, N. Hewitt, How heat pumps and thermal energy storage can be used to manage wind power: A study of Ireland, *Energy*, 157 (2018) 539-549.
- [12] D.L. Summerbell, D. Khripko, C. Barlow, J. Hesselbach, Cost and carbon reductions from industrial demand-side management: Study of potential savings at a cement plant, *Applied energy*, 197 (2017) 100-113.
- [13] P. Bradley, A. Coke, M. Leach, Financial incentive approaches for reducing peak electricity demand, experience from pilot trials with a UK energy provider, *Energy Policy*, 98 (2016) 108-120.
- [14] E.A.M. Ceseña, N. Good, P. Mancarella, Electrical network capacity support from demand side response: Techno-economic assessment of potential business cases for small commercial and residential end-users, *Energy Policy*, 82 (2015) 222-232.
- [15] P. Palensky, D. Dietrich, Demand side management: Demand response, intelligent energy systems, and smart loads, *IEEE transactions on industrial informatics*, 7 (2011) 381-388.
- [16] C.W. Gellings, *The smart grid: enabling energy efficiency and demand response*, River Publishers 2020.
- [17] T. Strasser, F. Andrén, J. Kathan, C. Cecati, C. Buccella, P. Siano, P. Leitao, G. Zhabelova, V. Vyatkin, P. Vrba, A review of architectures and concepts for intelligence in future electric energy systems, *IEEE Transactions on Industrial Electronics*, 62 (2014) 2424-2438.
- [18] P. Siano, Demand response and smart grids—A survey, *Renewable and sustainable energy reviews*, 30 (2014) 461-478.
- [19] J.C. Augusto, P. McCullagh, V. McClelland, J.-A. Walkden, Enhanced healthcare provision through assisted decision-making in a smart home environment, 2nd Workshop on Artificial Intelligence Techniques for Ambient Intelligence, 2007, pp. 27-32.
- [20] L. Tan, N. Wang, Future internet: The internet of things, 2010 3rd international conference on advanced computer theory and engineering (ICACTE), IEEE, 2010, pp. V5-376-V375-380.
- [21] R.J. Bessa, D. Rua, C. Abreu, P. Machado, J.R. Andrade, R. Pinto, C. Gonçalves, M. Reis, Data economy for prosumers in a smart grid ecosystem, *Proceedings of the Ninth International Conference on Future Energy Systems*, 2018, pp. 622-630.
- [22] C. Perera, A. Zaslavsky, P. Christen, D. Georgakopoulos, Ca4iot: Context awareness for internet of things, 2012 IEEE International Conference on Green Computing and Communications, IEEE, 2012, pp. 775-782.
- [23] S.A.U. Nambi, C. Sarkar, R.V. Prasad, A. Rahim, A unified semantic knowledge base for IoT, 2014 IEEE World Forum on Internet of Things (WF-IoT), IEEE, 2014, pp. 575-580.
- [24] V. Gazis, M. Goertz, M. Huber, A. Leonardi, K. Mathioudakis, A. Wiesmaier, F. Zeiger, Short paper: IoT: Challenges, projects, architectures, 2015 18th international conference on intelligence in next generation networks, IEEE, 2015, pp. 145-147.
- [25] I.D. Addo, S.I. Ahamed, S.S. Yau, A. Buduru, A reference architecture for improving security and privacy in internet of things applications, 2014 IEEE International conference on mobile services, IEEE, 2014, pp. 108-115.
- [26] B. Wu, T. Cheng, T.L. Yip, Y. Wang, Fuzzy logic based dynamic decision-making system for intelligent navigation strategy within inland traffic separation schemes, *Ocean Engineering*, 197 (2020) 106909.
- [27] R.S. Krishnan, E.G. Julie, Y.H. Robinson, S. Raja, R. Kumar, P.H. Thong, Fuzzy logic based smart irrigation system using internet of things, *Journal of Cleaner Production*, 252 (2020) 119902.
- [28] M. Yun, B. Yuxin, Research on the architecture and key technology of Internet of Things (IoT) applied on smart grid, 2010 international conference on advances in energy engineering, IEEE, 2010, pp. 69-72.
- [29] J. Tan, S.G. Koo, A survey of technologies in internet of things, 2014 IEEE International Conference on Distributed Computing in Sensor Systems, IEEE, 2014, pp. 269-274.
- [30] S. Bin, L. Yuan, W. Xiaoyi, Research on data mining models for the internet of things, 2010 International Conference on Image Analysis and Signal Processing, IEEE, 2010, pp. 127-132.
- [31] A. Garrab, A. Bouallegue, R. Bouallegue, An agent based fuzzy control for smart home energy management in smart grid environment, *International Journal of Renewable energy research*, 7 (2017) 599-612.
- [32] D.d.A. Ohi, N. Pereira, B.d.A. Prata, G. Barroso, Proposed architecture for energy efficiency and comfort optimization in smart homes, *Journal of Control, Automation and Electrical Systems*, 29 (2018) 718-730.
- [33] B. Asare-Bediako, W.L. Kling, P.F. Ribeiro, Multi-agent system architecture for smart home energy management and optimization, *IEEE PES ISGT Europe 2013*, IEEE, 2013, pp. 1-5.
- [34] Novák, V., Perfilieva, I. and Mockor, J., 2012. *Mathematical principles of fuzzy logic* (Vol. 517). Springer Science & Business Media.
- [35] Govindraj, Vignesh, Mithileysh Sathiyarayanan, and Babangida Abubakar. Customary homes to smart homes using Internet of Things (IoT) and mobile application. 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon). IEEE, 2017.
- [36] P. J. Rani, B. Jason, K. U. Praveen, K. U. Praveen, and K. Santhosh, Voice controlled home automation system using natural language processing (NLP) and internet of things (IoT),” in *Proceedings of the Third International Conference on Science Technology Engineering and Management*, IEEE, Chennai, India, March 2017.
- [37] H. Aadel and J. Ali, Design and implementation prototype of a smart house system at low cost and multi-functional, in *Future Technologies Conference (FTC)*, San Francisco, 2016.
- [38] Parsa, A., Najafabadi, T.A. and Salmasi, F.R., 2017, December. Implementation of smart optimal and automatic control of electrical home appliances (IoT). In *2017 Smart Grid Conference (SGC)* (pp. 1-6). IEEE
- [39] Xiaodong, Z. and Jie, Z., 2018, June. Design and implementation of smart home control system based on STM32. In *2018 Chinese control and decision conference (CCDC)* (pp. 3023-3027). IEEE.

# The Effect of Augmented Reality Mobile Application on Visitor Impact Mediated by Rational Hedonism: Evidence from Subak Museum

Ketut Agustini<sup>1</sup>, Dessy Seri Wahyuni<sup>2</sup>, I Nengah Eka Mertayasa<sup>3</sup>, Ni Made Ratminingsih<sup>4</sup>, Gede Ariadi<sup>5</sup>

Department of Informatics and Engineering Education-Faculty of Engineering and Vocational,  
Universitas Pendidikan Ganesha, Bali, Indonesia<sup>1,2,3</sup>

Department of English Language Education-Faculty of Language and Art, Universitas Pendidikan Ganesha, Bali, Indonesia<sup>4</sup>

Department of Management-Faculty of Economics and Business, Satya Wacana Christian University, Indonesia<sup>5</sup>

**Abstract**—This study expands our comprehension of museum visitor impact within a system quality, information quality, and augmented reality (AR) media content quality on mobile applications. Museums meet new defiance of escalating expectancies of their visitors. As a result of the universal mobile phone tool, AR has arisen as the latest technology offered to the museum to increase its visitors. These expectancies are fostered by the improvement of modern technologies like AR on the mobile app. Across an online survey of 241 visitors, the study determines the constructs affecting visitor impact within museum' mobile apps and the consequential results of AR-linked visitor impact. The study proposes a recent set of AR features, explicitly, system quality, information system, and AR media content quality, and establishes their influence on rational hedonism and satisfaction experienced, thus enhancing visitor impact. The findings also show that the rational hedonism and satisfaction experienced are positioned as full mediators for the relationship between system quality & information quality and visitor impact. In contrast, these mediators partially influence the indirect relationship between AR media content quality and visitor impact. Moreover, the results affirm that AR media content quality within the mobile application is the most critical construct to directly enhance visitor impact, whereas the system quality and information quality have no influence yet. From a practical point of view, the importance of AR technology for the museum can support entice new visitors to museums and improve to make more incomes.

**Keywords**—System quality; information quality; augmented reality media content quality; rational hedonism; satisfaction experienced

## I. INTRODUCTION

The importance of museums as artistic and heritage organizations is established, which have a substantial part in magnetizing new visitors to endpoint. To entice visitors and deliver them with good experiences, museums have begun to embrace modern technologies. Some scholars propose which are still significant alterations among museums based on high-tech tools, determined by the sum of aspects such as museums' size, site, income, and managerial control in where museums in emerging states are a significant burden to foster their technological capabilities [1][2].

Museums meet recent challenges of increasing the prospects of their guests. These outlooks are fostered by

improving the latest technologies such as virtual reality and augmented reality on mobile applications [3]. The escalation of information technology and digital forms altered how museums involved their visitors [2]. The improvement of digital media with mobile devices through augmented reality (AR) technology has gained a place in various users' hearts and many supporting applications [4]. AR has distinctive capability features to interlace the virtual and the material, implying museum visitors can suppress themselves in virtually improved experiences touching historical objects such artifacts, lontars, Subak (Balinese irrigation system) via AR apps. This study contends which AR media content provides creative potentials for assigning the unique features of historical objects, influencing satisfaction experienced and the enjoyment of the tour, like hedonism, and consequently increasing visitor impact [5].

Thus far, the recognition of AR mobile still meets several defiances explicitly for customer shopping engagement as the limited quality of media content and low quality of media layout[6]. Additionally, other concerns are linked to using the mobile application, such as lack of information quality and lack of system quality with mobile application because it is difficult to operate the function and content is not interactives [7][8]. These challenges can potentially increase visitors' resistance to using the AR mobile application. Conversely, there has been a lot to assess the influence of AR mobile's quality, system, and information on the acceptance of technology. System quality of the mobile application is essential to a beneficial user experience of mobile applications [9]. Determining the quality of AR mobile application is a multidimensional practice concentrating on several features of a system like system attributes, quality attributes, serviceability elements, and other aspects linked to technical matters [10]. System quality is the degree of easiness of utilizing and conveying out of assignments [11]. It also exhibits the significance of navigability, user-friendliness, organize, graphic logic, and stable mobile systems for enhancing a sound visitor capability and acquiring [12][13]. Some cosmetic, fashion and sports companies have implemented AR media to improve the authentic experience of their objects and assist customers through decision making [14]. Therefore, this paper tries to focus on the research gap by scrutinizing the impact of

the implementation of the AR mobile application on visitors in the Subak museum.

To fulfill the research gap, drawing from collaboration among the theory IS success model, the theory of Unified Theory of Acceptance and Use of Technology (UTAUT2) [13], and pure happiness theory [15][16], we construct and empirically emphasize a composition propositioning which, namely actual rational hedonism and satisfaction experienced respectively mediate the connection among system quality, information quality and AR media content quality, and visitor impact. The objective of this study is three aims. First, to investigate the constructs which affect visitor impact through AR attributes on mobile apps. Second, to evaluate the effect of rational hedonism to bridge the relationship between AR features and visitor impact. Thirdly, to assess the influence of satisfaction experienced as the mediator of the link between AR features and visitor impact.

## II. BASIC THEORY

### A. Factors Affecting Adoption of AR Mobile

Furthermore, IS success model draws information quality as an essential element to an information systems accomplishment [17]. A primary function of an AR mobile is to deliver visitors with figures appropriate to the historical objects [18]. Information, which is evident, recent, appropriate, precise, comprehensive, and consistent is trusted to be of excellent quality [19]. The performance of a technological system such as augmented reality is significant in instruction for consumers to implement and utilize such technology [20]. Then, a high-tech system is regarded as simple to utilize, which lets individuals finish duties, foster their output while also improving their performance and productivity [21]. In the context of this study, information quality and system quality extend to a visitor's view of the expectancies of significant others concerning the utilize of AR via mobile app.

Subsequently, UTAUT2 was initiated to comprise human behavior and practice [22]. The enjoyment reflects hedonism that has been considered a critical aspect of technology theory [22]. It is the action of expanding a particular system that is enjoyable in its own right, away from any performance effects ensuing from system use. The empirical study draws which enjoyment can affect mobile app utilization [20]. AR media content is imagined delivering a positive and performing experience [23]. Then, hedonism has converted into a fundamental construct in interpreting the adoption and utilization of the latest technology.

Regarding the pure happiness theory, the characteristic of a person's life is dependent on happiness, that is, the cognitive aspect, the hedonistic aspect, the mood aspect (emotional), and the hybrid aspect [15]. In this study, we try to implement the combination of cognitive and hedonistic views called rational hedonism. Each person is slightly hedonistic modestly because every person desires to experience an enjoyable life [24]. So, rational hedonism involves the pursuit of pleasure, a happy condition regarded on rational principles that create more significant comfort.

As a result, through integrating each theory (IS success model, UTAUT2, and pure happiness), the study encapsulates

technology, social and self-correlated constructs in the improvement of the theoretic model, which scrutinizes the part of rational hedonism and satisfaction experienced in enlightening visitors' degree of engagement by AR mobile software.

### B. Hypotheses Development

Particularly, virtual portrayals can influentially create the quality and historical objects via the mobile system; an improved configuration of a virtual object with visitor carries a new figure of exclusiveness; the cover of the virtual artifact-linked object on physical environments can generate object proximity and thus in-depth resonance by visitors than looking directly. The previous empiric has recognized how AR encourages sensorial elements and enhances individual performance [25][26]. Thus, the hypothesis is intended:

H1a: Higher system quality has a positive influence on visitor impact.

The magnitude of the system quality establishes a significant effect on the mobile application [27]. The system quality of mobile would interface which discovers new methods for collaboration between real and virtual surroundings simultaneously [28]. Additionally, it would enhance the attentiveness of the visitor. The accessible interface enables more artifact visualizations to their visitors, thus increasing the level of enjoyment. Such a platform enables the user to walk through the virtual museum and indicates the virtual depictions super-imposed on physical environments [29]. Correspondingly, Imaginings generation as augmented reality affordance permitting visitors to visualize 3D object portrayals in contrast to the actual condition [14].

Furthermore, system quality is acknowledged as easiness of learning, easiness of utilizing, the ease of accessibility, the effectiveness of system attributes, system complication, system attributes, and response time of IS [30]. It is as a degree of how well a supplied system meets visitor hopes such as effectual, reliable, adaptable, practical, customized, and fulfills the requirements [8]. Because visitors frequently settle in the comfortable system for longer time, the robust system makes enjoyment to consistently use the app, which aids encourage visitors' experiences [31]. The excellent system frequently causes visitor interfaces, which convey a perception of consistency, receptiveness, assurance, and compassion that can help visitors reach their aims and prospects more knowledgeably or, in this context, achieve hedonism [32]. Some empirical studies have revealed which people with these features positively generate pleasing and enjoyable experiences for visitors [33][32]. Then, the hypothesis is intended:

H1b: Higher system quality has a positive influence on rational hedonism.

System quality contains the required features (e.g., easiness of utilizing, flexible system, and reliable system) of mobile application [34]. Determining the quality of IS is a multi-aspect procedure converging on some attributes of a system like system attributes, quality attributes, usability attributes, and other attributes linked to technical concerns [35]. Characteristic determines system excellence in conventional studies involve response time and robustness [36]. AR mobile system

generates more revelational moments by indicating magnificence features in an immersive and sensory-rich approach. AR features create historical object components proximal and graphically high-up in physical environments, even though the scenery of the linked experiences varies [29]. Captivating this contextual museum, more excellent system quality is estimated to increase satisfaction experience. Thus, the hypothesis is composed:

H1c: Higher system quality has a positive influence on satisfaction experienced.

Information quality emphasizes the quality of a visitor's mobile's outcome (i.e., the information quality which the mobile delivers) and its helpfulness. It has been exhibited to be a critical success aspect while examining whole information system accomplishment [37]. Rich substance delivers the quality of the content concerning its practicality, comprehensibility, and robustness [38]. Some empirical studies have shown which information quality positively influences the individual impact [39][17][40]. Based on this discussion, information quality enhances the museum visitor impact.

Mobile informativeness positively influences purchase intentions, while information declines the degree of improbability to the products and enhances the possibility of shopping linked decision-making [41]. The information generated by the mobile system is the relative link and significance of a mobile regarded on consumer's requirements and concerns [42]. The mutual interaction and the quantity of information shared online affect cognitive contribution [43][44]. Similar to the museum context, the more informative and AR mobile, the more visitors endeavor to take back information that improves their participation [41]. Then, the following hypothesis is proposed:

H2a: Higher information quality has a positive effect on visitor impact.

The information quality portrays the exactness and precision of the information delivered by the mobile application; rightness is an alternative key indicator of information quality which information must be produced within proper and correct. So, visitors can get the latest adequacy data related to historical objects; adequacy is another attribute of information quality that should be appropriate and should comprise all information needed to the end-user [45]. Understandability is a valuable attribute of information quality, which must be simple to recognize and not be complicated and hard to comprehend. Conciseness is an attribute of information quality generated by the IS [46]. Moreover, information quality shows the outcome features of the mobile app's information, which corroborates the appropriate information toward museum visitors. The essential information is accessible at the exact time to the right man; the info delivered by the information mobile must be clear to the end-users [47].

Moreover, the spending of information related to history and culture can be inferred as a configuration of edutainment since visitors can acquire something and relish it simultaneously [48]. This is aligned with museums' aim of delivering knowledge and amusement, which relates with a dual method to recent visitor conduct where 'cognitive and

emotional aspects are concurrently at perform' [49]. It suggests that information generated by mobile enhances the level of psychological developments like enjoyment, hedonism, and affections [50]. The more helpful information generated by mobile applications provides the latest and relevant information; further, it is seen as informative, and visitors have more delight through the collaboration [41]. Thus, the hypothesizes that:

H2b: Higher information quality has a positive effect on rational hedonism.

Information quality can improve to enhance the satisfaction of the experienced visitor in the museum [51]. This empirical matches study from the perspective of museum visitors that exposed collaborative information affects the satisfaction experienced when visiting the museum [52]. Information quality plays a vital part in expanding positively to the advantages of handling a particular information technology (IT) [53]. It is viewed as the primary antecedent of user satisfaction [35]. Also, the increase of visitors' satisfaction due to it encourages them greater than static demonstrations. Visitors have more participated in the display while they can apply interactive displays [54]. Thus, the hypothesis is:

H2c: Higher information quality has a positive influence on satisfaction experience.

The existence of high-tech in museums turns out to be vital and must not be ignored. For instance, the revolution of museum experiences aims to the significance of interactive technology devices that can assist in accomplishing the preferred alters with the aid of explanatory texts and feasible activities [55]. Consistently, which museum experts believe AR technology is a museum instrument whose incorporation into museums generates the foundation for interface via involved entertainment [56]. Using multimedia features such as AR with interactive content can positively affect learning development [57]. Similar to the museum context, virtual object generated by AR app enhances learning progression by sustaining the visitors' engagement. For instance, the implementation of videos and virtual animations can enlighten complicated concepts more effectively rather than concern the text [57]. Regarded above this discussion, the AR media content quality corroboration aspect can act as a vital aspect in leveraging the museum visitor impact. Thus, the hypothesis is:

H3a: High AR media content quality has a positive influence on visitor impact.

AR is the incorporation of digital data with the user's situation. It utilizes the present surrounding and overlays the latest information on top. This technology enhances the digital creation with physical targets, letting a real-domain customer flawlessly interrelate with digital elements [58]. The AR utilization in an art museum has a vital role in the recent improvements linked to acquiring technology that considers the latest high-tech instruments as a proposal to encourage the learning process [58]. Furthermore, AR is present converting human developments by speeding up competencies progress and encouraging guidance.

The amazement of AR is escalating because it carries components of the virtual object into the actual environment,

then augmenting the objects for seen, heard, and perceived. It conveys computer-created objects into the reality surrounding, but the visitor can look at the historical object. When utilizing AR apps, the user imagines a mix of synthetic and natural brightness [59]. Overlaying pictures are projected above historical objects, which allocates the figures and collaborative virtual objects to be on top of the user's interpretation of the natural environment. AR can be very proper for replications, specifically in the museum zone [51].

Furthermore, AR technology integrates aspects of amusement and collaboration into museum propose to entice new visitors, particularly the children and teenager segment[60]. For instance, the Art Museum in the USA utilizes the influence of technology to involve visitors enthusiastically. With the aid of the AR app, visitors can perform a mainly intended museum game. They are encouraged to reconstruct the stances of the sculptures in the museum's gallery. Then they are delivered by information related to the art they are attempting to reconstruct in the game. In this approach, a visitor has amusing and learns together with the support of collaborative technology. By embracing AR media content, interactive installments entice visitors and create them more dynamic for an extended time than standard static displays [54]. Collaborative link technologies create the apparent authenticity of visitors' familiarity in a museum [50]. The museum embraces content entertainment through displays that foster all visitors to absorb more about themselves and others while they go to the museum [51]. So, technologies in museums (AR) can induce visitors' emotional reactions such as enjoyable and hedonism [50][6]. Thus:

H3b: High AR media content quality has a positive influence on rational hedonism.

AR technology delivers objects to improve the visitor tour by expanding magnificence features animatedly through diverse touchpoints crossways the tour itself [29]. AR's virtuality authentically replicates and increases the more excellent quality of objects while covered in the natural realm or lined up with visitors' looks. Moreover, objects can implement AR to certify the easy extension of the tour via facilitating fewer jarring alterations between virtual and physical touchpoints. This study highlights how AR content aids historical objects in delivering high-end facilities through zooming the high resolution for their objects, which satisfaction [61]. So, AR generates more revolution moments by demonstrating magnificent content in an immersive and sensory-rich approach. Its features generate object media proximal and graphically prominent in physical environments, so that increase satisfaction experience. Thus, the following hypothesis is proposed:

H3c: High AR media content quality has a positive effect on the satisfaction experience.

Regarding the pure happiness theory, the characteristic of a person's life is dependent on happiness, that is, the cognitive aspect, the hedonistic aspect, the mood aspect (emotional), and the hybrid aspect [15]. In this study, we try to implement the combination of cognitive and hedonistic views called rational

hedonism. Cognitively evaluating one's life as an entire in a positive approach makes to perceive rationality. Each person is slightly hedonistic modestly because every person desires to experience an enjoyable life [24]. So, rational hedonism involves the pursuit of pleasure simply, which is a happy condition regarded on rational principles that creates more excellent support to comfort.

Rational hedonism is linked with a fun and exciting experience which may ensue in enjoyment. An enjoyable and fun mobile application creates positive hedonic contagion by augmented reality (AR) technology [62]. Since augmented reality (AR) media implements the natural surrounding and overlays recent information above it which appears to augment the digital domain with physical objects, letting a real-world visitor flawlessly interrelate with digital elements [59]. It serves users to avoid tediousness to rest; hedonism can be a vital feature to generate unique sensory emotion [29]. In entirety, hedonic understandings create changes in the users' emotional testifies, making a positive response that enhances their aims to use AR mobile.

Hedonism distributes with the customers' affections (happy and adoring) and has arisen as a vital aspect in influencing social behavior [63][64]. The joyful and ludic part of visiting the shop is that visitors are more involved with the practical part (fun) of relaxation [65]. In the museum context, visiting emphases are considered the hedonic value [66]. Visitors appreciate obtaining engaged in the touring process and are more attracted to the fun aspect of the touring experience [67]. Scholars have discovered which faithfulness of AR app improves if the visitors are satisfied with the visual involvements, i.e., degree of enjoyment [64]. Hedonism has been realized to affect the satisfied visitor positively, then encourage the visitor to revisit [68][6]. Then, the hypothesis is proposed:

H4: Rational hedonism has a positive influence on visitor impact.

The satisfaction experience has a substantial effect on behavioral intentions [69]. The feelings customers create are bound to the aptness of the services obtained (i.e., the contrast between performance and hopes [70]). They construct potential developments of engagements on their level of satisfaction [71]. Satisfaction with a product or service is the essential enthusiasm to persist applying it as satisfaction indicates which advantages originated from visualizing media content by AR application. In the museum context, the satisfaction experience with AR mobile would indicate the achievement of the estimated advantages, improving the visitor impact to use this application to maintain acquiring these advantages [72]. In the AR mobile context, the visitor impact enables acquiring local knowledge "Subak"; to improve the exploring historical object; to makes easier using the application; to improve technological skills. The increased visitor satisfaction experience will consequently influence visitor impact. Then, the hypotheses is:

H5: High satisfaction experienced has a positive effect on visitor impact.

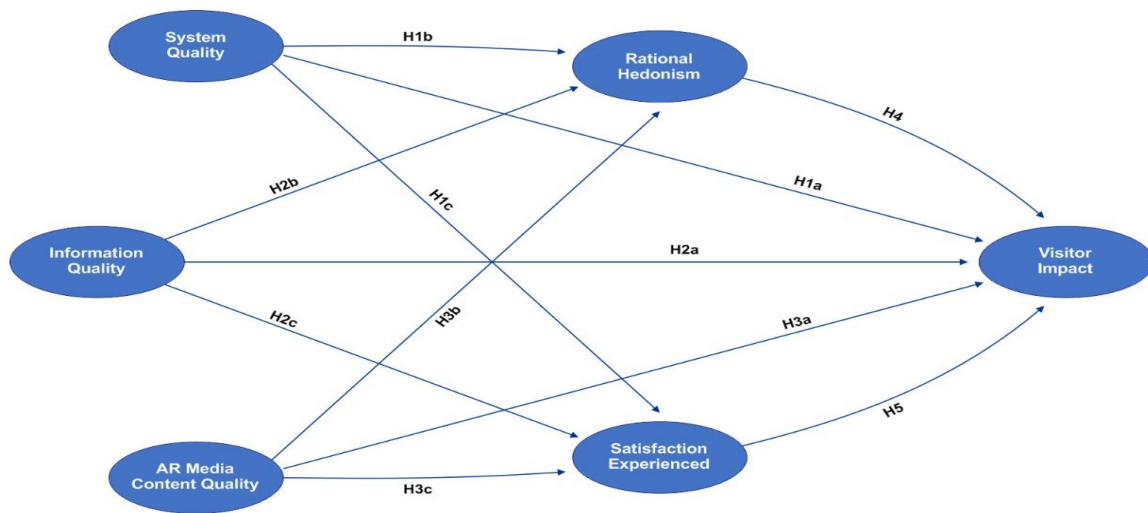


Fig. 1. Conceptual model.

Based on the above arguments, the satisfaction experienced, and the rational hedonism has a significant role in mediating the total effect of system quality, information quality, and AR media content quality on visitor impact. Thus, the hypotheses proposed are (Fig. 1):

H6a: The rational hedonism mediates the impact of system quality on visitor impact.

H6b: The rational hedonism mediates the impact of information quality on visitor impact.

H6c: The rational hedonism mediates the impact of AR media content quality on visitor impact.

H7a: The satisfaction experienced mediates the impact of system quality on visitor impact.

H7b: The satisfaction experienced mediates the impact of information quality on visitor impact.

H7c: The satisfaction experienced mediates the impact of AR media content quality on visitor impact.

### III. METHODOLOGY

#### A. Data Collection and Sample

The research employed a quantitative approach. Collected information on the critical research was obtained by the visitors of the Subak Museum at Bali Province, Indonesia. The data were gathered employing a survey handled through google form between May 2021 and July 2021. All items were calculated using five-point Likert scales, varying from totally disagree (1) to totally agree (5). A pilot study was performed on a group of 30 museum visitors who were not involved in the primary survey to examine the instrument. An overall of 427 e-mails was disseminated in May 2021, specifying the hyperlink to the survey and requesting its contribution. Two hundred seventy-nine responses were obtained toward the end of July 2021, which relates to a 65.33 percent response rate. Thirty-eight responses were dropped due to inadequate, resting 241 (56.44 percent) with valid and finished responses. In Table I, the resulting data were gathered from 241 visitors that visited Subak Museum in Indonesia.

TABLE I. DEMOGRAPHIC PROFILE OF RESPONDENTS (N=241)

Demographics Respondent Percentage (%)	Frequency	Percentage
<i>Gender</i>		
Male	140	59,29%
Female	101	40,71%
<i>Age-Group</i>		
17-30 years	137	56,85%
31-40 years	86	35,68%
41 years above	18	7,47%
<i>Education</i>		
High School or below	104	43,15%
Undergraduate degree	87	36,09%
Bachelor	37	15,35%
Master's degree or higher	13	5,39%

#### B. Profile of Subak Museum

The Subak Museum is a museum that displays a collection of traditional agriculture tools utilized in farming and the traditional Balinese irrigation termed as Subak. It is one of the tourist objects for study tours, which many students and travelers from domestic and foreign tourists visit. Subak Museum is one of Bali's icons that stores agricultural tools of Balinese cultural heritage through the concept of Subak network has been recognized by the World Health Organization (WHO) as a World Cultural heritage. Museum is located in Banjar Anyar Village, Kediri District, Tabanan Regency, Bali - Indonesia. The distance from Denpasar City is approximately 21 KM. The number of visitors to Subak Museum from 2016 to 2020 decreased from 28,355 to 21,104 visitors.

History of the Subak Museum was established on August 17, 1975, to preserve the Subak traditional institution

as a noble cultural heritage of the nation and introduce the younger generation and tourists to the unique traditional irrigation system in Bali. Subak is an organization of farming communities in Bali that deals explicitly with the regulation or system of irrigating rice fields/irrigation in a traditional/conventional way.

The Subak Museum provides an experience for visitors to see, recognize and learn all things related to agriculture and traditional equipment commonly used in agriculture in Bali, such as rice cutting and pounding tools, rice plowing tools, etc. Furthermore, here visitors get complete information about how the Subak irrigation system and processing rice fields from start to finish. Such as how to open rice fields, share the water, make water tunnels, measure waterways, and complete the process of religious ritual ceremonies.

**C. Instrument Development**

Based on previous studies, five constructs have been established by the author for supporting student performance (System Quality; Information Quality; AR media content Quality; Rational Hedonism; Satisfaction experienced; Individual Impact). Four items determined system Quality: AR mobile is simple to navigate; easy to use; admits visitors to search the information; offers proper functionality [73]. Information Quality was measured by four items that information provided by AR mobile is practical, understandable, attractive, reliable [7]. Four items measured AR media content quality: AR mobile desires enhancement of object reality; delivers powerful momentary; creates captivating aesthetic simulations; create real-time interactivity between virtual and physical elements [74][75][29]. Rational hedonism was measured by four items that visitor enjoys their tour through the AR application; have incredible emotion through the AR technology; have exciting times to explore the historical object; have a discovery of local knowledge [76]. The Satisfaction experienced was measured by four items that the visitors feel satisfied with their experience; the experience matches what visitors expect, the experience has succeeded as well as visitor contemplates, the getting auratic from the visual object [6][29]. The visitor impact was measured by four items that museum visitors enable to acquire local knowledge, “Subak”, to improve the exploring historical object; to enhance technological skills; to make easier using the application [7]. The questionnaire was obtained from empirical studies which analyzed the variables in an inquiry in the current research.

**IV. DATA ANALYSIS AND RESULTS**

To recognize the direct and indirect effect of system quality, information quality, and AR media content quality on visitor impact with the mediating impact of rational hedonism and satisfaction experienced at museum visitors in Indonesia. For evaluating the entire measurement model, statistical SmartPLS software was applied, and analyzing for data was utilized by PLS-SEM (Partial Least Square-Structural Equation Modelling).

**A. Measurement Model**

Convergent validity and discriminant validity have been examined. We have evaluated convergent validity by examining factor loading greater than 0.7, composite

reliabilities greater than 0.8, and the average extracted variance (AVE) must greater than 0.5 for all variables (Fornell and Larcker, 1981). All factor charges in our model are greater than 0.7 and measuring objects are removed if their factor loads are less than 0.70. The finding results show that our model meets the standard of convergent validity. With Cronbach  $\alpha$ , we examined the internal reliability of scales. Table II states the loading factor, AVE, CR and (C- $\alpha$ ) of all constructs.

The latest approach was applied to confirm discriminant validity in the Heterotrait-Monotrait (HTMT) formula, and the HTMT values are displayed in Table III. If the HTMT is greater than 0.90, then the test of discriminant validity is the failure[77]. As all the results of HTMT are lower than the threshold portrayed in Table III, discriminant validity has been confirmed [77]. The measurement of the goodness-of-fit model was revealed to be satisfactory (Standardized Root Mean Square Residual [SRMR]=0.071, and Normal Fit Index [NFI]=0.928) and established the proposed model because of SRMR value < 0.08 and NFI value > 0.9 [78]. Conclusively, we postulate that tests of convergent and discriminant are valid, thus satisfactory to examine the hypothesis for the study.

TABLE II. CONVERGENT VALIDITY

Constructs	Items	Factor Loadings	Cronbach $\alpha$	Composite Reliability	AVE
System Quality (SQ)	SQ1	0,897	0,851	0,899	0,690
	SQ2	0,736			
	SQ3	0,801			
	SQ4	0,879			
Information Quality (IQ)	IQ1	0,885	0,722	0,770	0,703
	IQ2	0,712			
	IQ3	0,814			
	IQ4	0,853			
AR Media Content Quality (AR)	AR1	0,689	0,778	0,834	0,605
	AR2	0,854			
	AR3	0,766			
	AR4	0,787			
Rational Hedonism (RH)	RH1	0,908	0,725	0,836	0,645
	RH2	0,845			
	RH3	0,649			
	RH4	0,826			
Satisfaction Experienced (SE)	SE1	0,893	0,802	0,884	0,717
	SE2	0,812			
	SE3	0,834			
	SE4	0,745			
Visitor Impact (VI)	VI1	0,863	0,766	0,864	0,679
	VI2	0,758			
	VI3	0,848			
	VI4	0,793			

TABLE III. DISCRIMINANT VALIDITY

	SQ	IQ	AR	RH	SE	VI
SQ						
IQ	0,455					
AR	0,284	0,701				
RH	0,509	0,272	0,408			
SE	0,349	0,798	0,759	0,684		
VI	0,509	0,812	0,743	0,351	0,682	

**B. Hypothesis Examining**

The study examines the structural relationships among the variables by investigating the diverse mediation effects across a path analysis. Path analyses were applied to test hypotheses in the conceptual model employing the Smart-PLS software. The research shown in Table IV depicts path coefficients of the study model. Table IV and Fig. 2 describes which the path values from System Quality to Visitor Impact was positive and non-significant ( $\beta = 0.121$ ;  $p > 0.1$ ), in contrary the path values from System Quality to Rational Hedonism was also positive and significant ( $\beta = 0.544$ ;  $p < 0.01$ ) and the path coefficients from System Quality to Satisfaction Experienced was also positive and significant ( $\beta = 0.278$ ;  $p < 0.05$ ). Therefore,  $H_{1a}$  is not supported but  $H_{1b}$  dan  $H_{1c}$  are supported. The path coefficient from Information Quality to Visitor Impact was positive and non-significant ( $\beta = 0.095$ ;  $p > 0.1$ ), whereas the path coefficient from Information Quality to Rational Hedonism was positive and significant ( $\beta = 0.162$ ;  $p < 0.1$ ), and the path coefficient from Information Quality to Satisfaction Experienced was positive and significant ( $\beta = 0.145$ ;  $p < 0.1$ ). Then, there is enough evidence to support  $H_{2b}$  and  $H_{2c}$  but  $H_{2a}$  was not supported. The path coefficient from AR Media Content Quality to Visitor Impact was positive and significant ( $\beta = 0.257$ ;  $p < 0.05$ ), and the path coefficient from

AR Media Content Quality to Rational Hedonism was positive and significant ( $\beta = 0.201$ ;  $p < 0.05$ ), also the path coefficient from AR Media Content Quality to Satisfaction Experienced was positive and significant ( $\beta = 0.473$ ;  $p < 0.01$ ). Thus, there is enough evidence to support  $H_{3a}$ ,  $H_{3b}$  and  $H_{3c}$ . The path coefficient from Rational Hedonism to Visitor Impact was positive and significant ( $\beta = 0.295$ ;  $p < 0.01$ ), and the path coefficient from Satisfaction Experienced to Visit or Impact was positive and significant ( $\beta = 0.254$ ;  $p < 0.05$ ). Therefore,  $H_4$  and  $H_5$  are supported.

Additionally, the indirect relationship of System Quality, Information Quality, and AR Media Content Quality on Visitor Impact through Rational Hedonism as mediator was also positive and significant respectively ( $\beta = 0.161$ ,  $p < 0.05$ ;  $\beta = 0.048$ ,  $p < 0.1$ ;  $\beta = 0.059$ ,  $p < 0.1$ ), that  $H_{6a}$ ,  $H_{6b}$ , and  $H_{6c}$  are supported. Then, the indirect effects of System Quality, Information Quality, and AR Media Content Quality on Visitor Impact through Satisfaction Experienced as mediator was also positive and significant ( $\beta = 0.071$ ,  $p < 0.1$ ;  $\beta = 0.036$ ,  $p < 0.1$ ;  $\beta = 0.12$ ,  $p < 0.05$ ), that state  $H_{7a}$ ,  $H_{7b}$ , and  $H_{7c}$  are confirmed. Based on the explanation above, we determine that Rational Hedonism and Satisfaction Experienced partially mediate the relationship between System Quality, Information Quality, and AR Media Content Quality and Visitor Impact.

TABLE IV. HYPOTHESES TESTING

Hypothesis	Relationship	Standard Coefficients	Test Result
$H_{1a}$	System Quality → Visitor Impact	0,121	Non-Significant
$H_{1b}$	System Quality → Rational Hedonism	0,544*	Significant
$H_{1c}$	System Quality → Satisfaction Experienced	0,278 **	Significant
$H_{2a}$	Information Quality → Visitor Impact	0,095	Non-Significant
$H_{2b}$	Information Quality → Rational Hedonism	0,162***	Significant
$H_{2c}$	Information Quality → Satisfaction Experienced	0,145***	Significant
$H_{3a}$	AR Media Content Quality → Visitor Impact	0,257 **	Significant
$H_{3b}$	AR Media Content Quality → Rational Hedonism	0,201 **	Significant
$H_{3c}$	AR Media Content Quality → Satisfaction Experienced	0,473 *	Significant
$H_4$	Rational Hedonism → Visitor Impact	0,295 *	Significant
$H_5$	Satisfaction Experienced → Visitor Impact	0,254 **	Significant
$H_{6a}$	System Quality → Rational Hedonism → Visitor Impact	0,161 **	Significant
$H_{6b}$	Information Quality → Rational Hedonism → Visitor Impact	0,048***	Significant
$H_{6c}$	AR Media Content Quality → Rational Hedonism → Visitor Impact	0,059 ***	Significant
$H_{7a}$	System Quality → Satisfaction Experienced → Visitor Impact	0,071 ***	Significant
$H_{7b}$	Information Quality → Satisfaction Experienced → Visitor Impact	0,036***	Significant
$H_{7c}$	AR Media Content Quality → Satisfaction Experienced → Visitor Impact	0,12**	Significant

Note: Significant at \*1%, \*\*5% and \*\*\*10% levels



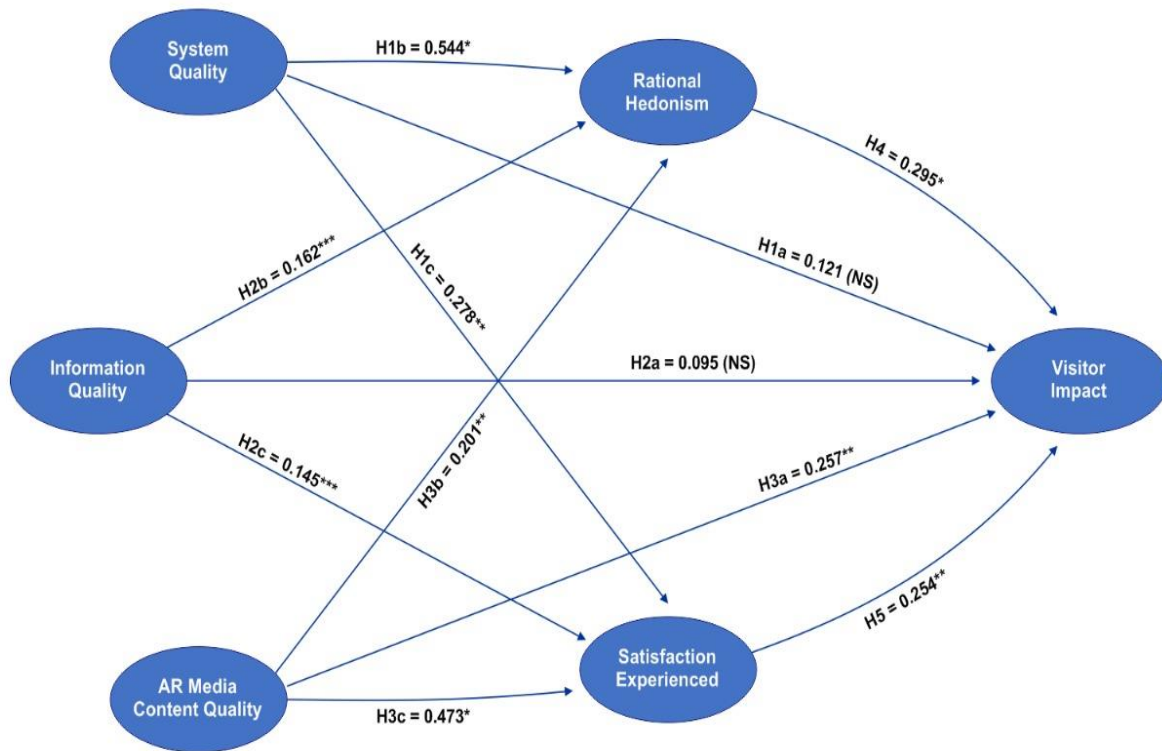


Fig. 2. Results of Path Analysis.

## V. DISCUSSION

The current study results establish which the system quality does not influence the visitors' impact in Subak museum, which corroborates the findings [79]. Some museum visitors have difficulty finding information contained in AR mobile, and some of the existing features are still unstable. Concurrently, several visitors suppose technological constraints often obtained in new, inventive technologies, such as arduousness in installing the related software, deficiency of computer literacy, and malfunctioning of the AR applications (e.g., slow response speediness, cartoony object pictures). Then, AR is assumed to be at the growing phase, requiring extra scope to enhance to be adored and utilized by more visitors. On the other hand, another empirical study reveals which system quality has a significant effect on rational hedonism, which aligns with the findings [80][81]. Therefore, the system quality is an essential aspect for hedonic-seeking visitors while they involve in openness services. Some visitors obtain it central for the system delivered to be stable with their hopes. Then, AR applications should endeavor to recognize their visitors' desires and support them in accomplishing their aims also is expected to provide a playful and entertaining experience for visitors. Moreover, the other finding shows that system quality has a significant effect on satisfying experiences, which aligns with the findings [6][79][8]. The system technology's ability enables visitors to cooperate more, handle and suit immersed with objects. Possibly, AR delivers visitors with the most interactive kind of technology accessible, where persons have managed over their manipulation in

joining both the actual object and virtual object environment [82]. The AR capability can create mental figures which reproduce historical objects and experiences, which is the main competency through visitors observing the artifact. Visitors have long tried to visualize the objects at the museum to understand their applicability before looking at the actual objects.

Conversely, the other empirical finding shows that the information quality does not influence visitor impacts in Subak museum, which aligns with [7]. The display of information on the application screen is less reliable with the information in the physical diorama, so that there is a misinterpretation for museum visitors. For this reason, visitors could not fully capture the explanation given from the application, such as information about artifacts with precision. In contrast, another empirical study reveals that information quality positively affects rational hedonism, which aligns with the findings [83][80]. The information generated by AR apps enables visitors to personalize content to their personal preferences and fascinations. Moreover, AR applications could deliver further historical objects information like supplementary video media and text matter comprising the process of Subak irrigation and material information of artifacts. The function that originates from this unique subject is possible to improve a visitor's enjoyment by having exciting times to explore the historical object, obtaining local knowledge, and getting incredible emotion. The following finding shows that the information quality has a positive effect on satisfaction experienced, which aligns with the findings [84][72]. Information is evident, recent, appropriate, precise, comprehensive, and consistent and

is of excellent quality [85]. This paper shows which higher-quality information is a vital aspect in attaining satisfaction with AR mobile. It offers visitors information, which is recent, wide-ranging, simply explicable, and converges the information required to enhance a visitor's degree of satisfaction. Information could air to improve the sensed realism of the museum experienced visitors. They expect AR mobile to offer capable and reliable information which would satisfy their information necessities. This result match by study from the viewpoint of museum visitors that revealed interactive information significantly influences the sensed realism of the museum [50]. So, visitors displayed with more excellent informative object visualizations (i.e., artifacts) indicate a more positive experience.

Furthermore, the other finding demonstrates that the AR media content quality influences visitor impacts in Subak museum, which aligns with the findings [2]. AR technology generates and provides content in an approach that visitors favor using. It could enhance the sensed realism of material objects, which supports museums in describing and interactively showcasing the objects. For example, the Subak Museum applies museum's high-tech facility intensifies with praxis, its capability to transform and explain collections subject improves to generate attractive and impressive museum experiences [82]. Moreover, another empirical study shows that the AR media content quality positively impacts rational hedonism, which aligns with the findings [29][6]. AR content could support improving the sensed enjoyment of the visitor because it involves them more than motionless expositions. Visitors have more participated in the expositions while they have a chance to utilize interactive displays. Virtual reality overlays on or is incorporated as a role, and physical environments can graphically point out the delicate materials and artifacts. The visitor has the facility to improve the magnificence aura by envisioning object heritage. Application authentically replicates and strengthens the more outstanding quality of historical objects while covered in the natural environment. The development of supporting virtual finest object materials with visitors' emergence attributes is central for fitting visitor necessities and customizing the object trial. Mainly, virtual depictions could realistically imitate the quality and resolution of historical heritages in a way not previously likely through a website; the overlay of virtual object-related content on physical surroundings can create object proximity and, therefore, a deeper resonance with visitors than online browsing. The following finding shows that the AR media content quality positively affects satisfaction experienced, which aligns with the findings [29][6]. Regarding museum experience, this research identifies that AR media content provides an attentiveness of historical object stimuli round detailed dioramas. Specifically, it allows historical objects to induce influential emotions in visitors via AR mobile, short-term faces which are intensively satisfied due to the high-quality resolution and unique content for particular objects. Conversely, auratic magnification provides a new approach to constructing an object's distinctive aura – for example, by envisioning the object's realism artifact and distinction in an immersive method. For essentially encouraged museum visitors, the acquiring experience might be improved by AR media content, which becomes a crucial aspect of the visitor's

experience. It provides visitors with information, which is up-to-date, complete, simply comprehensible, and matches the information requirement enhances a visitor's degree of satisfaction. So, AR technology helps museums obtain attractiveness by facilitating visitors to achieve experiences that are concurrently aesthetically educational and entertainment [86].

The following finding shows that rational hedonism has a positive effect on visitor impact, which aligns with the findings [21][6]. Hence, enjoyment might be caused due to the inventive visualization experience delivered by AR as visitors can manipulate objects and have the possibility to generate themselves a customizable experience. During a customer's experience, enjoyment can be obtained by simulating 3-D virtual objects [18][87]. Thus, a diversity of AR media attributes that deliver interaction and lucidity to visitors via AR might be adept at elevating a visitor's imaginative composition progression through their experience, incorporating their natural surroundings with the virtual environment to acquire a fun experience within the visualization of new objects. The other results reveal that the satisfaction experienced positively influences visitor impact, which aligns with the findings [80][72]. Museum visitors turn into satisfied, delighted, and relaxed with their judgments to utilize in AR mobile while the application could meet their necessities and facility offerings match their hopes. Furthermore, visitors feel a positive appraise while they trust that they have obtained reasonable offers for AR attributes. The visitor felt value could be determined by comparing what they give up (e.g., financial, time, power, and affection) with what they obtain (i.e., auratic experiences when looking at virtual objects). The satisfaction experienced by the visitor is obtained by quality of the museum experience, as well as the level of visitors fascinate, causes to overall museum visitors' satisfaction [88]. The opinion of visitors' satisfaction improves while the worth incorporation progression outcomes in a positive result are naturally more likely to foster positive effect.

## VI. THEORETICAL AND MANAGERIAL IMPLICATIONS

Theoretically, this study cooperates the collaboration among the theory IS success model, the theory of Unified Theory of Acceptance and Use of Technology (UTAUT2), and theory of pure happiness, also information quality, system quality, AR media content quality to increase visitor impact through rational hedonism, and satisfaction experienced in Subak Museum. The rational hedonism and satisfaction experienced are positioned as full mediators for the relationship between system quality and information quality and visitor impact. In contrast, these mediators partially influence the indirect relationship between AR media content quality and visitor impact. This exemplifies the new model, which fulfills the research gap among information quality, system quality, and AR media content quality on visitor impact. Additionally, the results affirm that AR media content quality within the mobile application is the most critical construct to directly enhance visitor impact, whereas the system quality and information quality have no influence yet.

From a practical point of view, the paper influences the stream of exploring the magnitude of AR technology for

museums. Introducing AR media content for museum delivery (offline & online) can support entice new visitors coming to museums, and therefore improve make more incomes. In achieving the interactivity of the visitor experience, the museum manager needs to recognize how the media content might be employed in the museum's proposal. The level of satisfaction experienced, and rational hedonism is summarized as a crucial aspect in influencing visitor impact. Then, managers must recognize the vital function AR content performs in improving visitor interactivity as the involvement of their getting satisfaction experienced from utilizing AR app. Lastly, the robust system quality of AR mobile creates perceived enjoyment that is termed as rational hedonism to increase visitor impact.

## VII. CONCLUSION AND LIMITATIONS

These empirical outcomes have provided some substantially beneficial shreds of proof of the role of system quality, information quality, and AR media content quality to increase visitor impact in the Subak Museum in Indonesia through rational hedonism and satisfaction experienced. The satisfaction experienced is more influential in the linkage between AR media content quality and visitor impact rather than system quality, and information quality as antecedents, whereas rational hedonism is more influential in the linkage between quality system and visitor impact. Thus, the study suggests which rational hedonism and satisfaction experienced fully mediate the relationship between system quality & information quality and visitor impact. Finally, the indirect relationship between AR media content quality and visitor impact is only bridged by rational hedonism and satisfaction experienced.

There are constraints that propose some future research. This analysis employs a cross-sectional design that will let a longitudinal study be applied for the following research to scrutinize the influences of features AR mobile application on rational hedonism and satisfaction experienced that also increases visitor impact. Lastly, this study is only conveyed by a single type of museum, and it is attractive lead to collect data from other museums to provide more corroboration of results.

## ACKNOWLEDGMENT

The authors would like to thank for the Directorate General of Research and Development, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia which had provided funding for this research through research grant number: 1219/UN48.16/LT/2022. The authors also would like to thank for Rector and Chair of the Research and Community Service Institute of Universitas Pendidikan Ganesha who provided the opportunity, guidance, and support to complete this research on time.

## REFERENCES

- [1] Y. Evrard and A. Krebs, "The authenticity of the museum experience in the digital age: the case of the Louvre," *J. Cult. Econ.*, vol. 42, no. 3, pp. 353–363, 2018, doi: 10.1007/s10824-017-9309-x.
- [2] M. C. Wang and J. Q.-P. Lin, "The Future Museum shapes the museum future," *Arts Mark.*, vol. 8, no. 2, pp. 168–181, 2018, doi: 10.1108/aam-12-2017-0030.
- [3] M. Carrozzino and M. Bergamasco, "Beyond virtual museums: Experiencing immersive virtual reality in real museums," *J. Cult. Herit.*, vol. 11, no. 4, 2010, doi: 10.1016/j.culher.2010.04.001.
- [4] M. J. Rodrigues et al., "Dataset on functional and chemical properties of the medicinal halophyte *Polygonum maritimum* L. under greenhouse cultivation," *Data Br.*, vol. 25, 2019, doi: 10.1016/j.dib.2019.104357.
- [5] C. Kuehnl, D. Jozic, and C. Homburg, "Effective customer journey design: consumers' conception, measurement, and consequences," *J. Acad. Mark. Sci.*, vol. 47, no. 3, 2019, doi: 10.1007/s11747-018-00625-7.
- [6] G. McLean and A. Wilson, "Shopping in the digital world: Examining customer engagement through augmented reality mobile applications," *Comput. Human Behav.*, vol. 101, 2019, doi: 10.1016/j.chb.2019.07.002.
- [7] W. A. Cidral, T. Oliveira, M. Di Felice, and M. Aparicio, "E-learning success determinants: Brazilian empirical study," *Comput. Educ.*, vol. 122, 2018, doi: 10.1016/j.compedu.2017.12.001.
- [8] A. Shahzad, R. Hassan, A. Y. Aremu, A. Hussain, and R. N. Lodhi, "Effects of COVID-19 in E-learning on higher education institution students: the group comparison between male and female," *Qual. Quant.*, vol. 55, no. 3, 2021, doi: 10.1007/s11135-020-01028-z.
- [9] C. Tam and T. Oliveira, "Understanding the impact of m-banking on individual performance: DeLone & McLean and TTF perspective," *Comput. Human Behav.*, vol. 61, 2016, doi: 10.1016/j.chb.2016.03.016.
- [10] Z. Zaremohzabieh et al., "A Model of Information Systems Success for Assessing the Effectiveness of Statistical Learning Tool on University Students Performance in Statistics," *Mediterr. J. Soc. Sci.*, 2016, doi: 10.5901/mjss.2016.v7n2p271.
- [11] A. M. Elkaseh, K. W. Wong, and C. C. Fung, "Perceived Ease of Use and Perceived Usefulness of Social Media for e-Learning in Libyan Higher Education: A Structural Equation Modeling Analysis," *Int. J. Inf. Educ. Technol.*, vol. 6, no. 3, pp. 192–199, 2016, doi: 10.7763/ijiet.2016.v6.683.
- [12] M. A. Butzke, A. Alberton, J. Visentainer, S. Garcia, and I. de Alencar Nääs, "Business games based on simulation and decision-making in logistics processes," in *IFIP Advances in Information and Communication Technology*, 2017, vol. 514, doi: 10.1007/978-3-319-66926-7\_11.
- [13] A. Tarhini, R. M. Deh, K. A. Al-Busaidi, A. B. Mohammed, and M. Maqableh, "Factors influencing students' adoption of e-learning: A structural equation modeling approach," *J. Int. Educ. Bus.*, vol. 10, no. 2, 2017, doi: 10.1108/JIEB-09-2016-0032.
- [14] M. Heller, A. M. Thomas, J. D. Klausner, S. M. Peters, and K. M. Düsterwald, "An Evaluation of Patient and Student Experience at a Longstanding Student-run Free Clinic in Cape Town, South Africa," *Cureus*, 2019, doi: 10.7759/cureus.6320.
- [15] B. Brülde, "Happiness theories of the good life," *J. Happiness Stud.*, vol. 8, no. 1, 2007, doi: 10.1007/s10902-006-9003-8.
- [16] O. Güler and M. İ. Haseki, "Positive Psychological Impacts of Cooking During the COVID-19 Lockdown Period: A Qualitative Study," *Front. Psychol.*, vol. 12, 2021, doi: 10.3389/fpsyg.2021.635957.
- [17] M. W. Makokha and D. O. Ochieng, "Assessing the Success of ICT's from a User Perspective: Case Study of Coffee Research Foundation, Kenya," *J. Manag. Strateg.*, vol. 5, no. 4, 2014, doi: 10.5430/jms.v5n4p46.
- [18] R. C. King and S. Dong, "The impact of smartphone on young adults," *Bus. Manag. Rev.*, vol. 8, no. 4, 2017.
- [19] X. Chen, X. Xiong, M. Zhang, and W. Li, "Public authority control strategy for opinion evolution in social networks," *Chaos*, vol. 26, no. 8, 2016, doi: 10.1063/1.4960121.
- [20] C. Y. Huang, C. J. Chen, Y. F. Lee, H. C. Yeh, J. C. Kuo, and H. L. Lai, "Effects of individual characteristics on insomnia severity trajectory among nurses: A prospective longitudinal study," *J. Nurs. Manag.*, vol. 27, no. 8, pp. 1640–1647, 2019, doi: 10.1111/jonm.12851.
- [21] M. K. Al-Kofahi, H. Hassan, and R. Mohamad, "Information systems success model: A review of literature," *International Journal of Innovation, Creativity and Change*, vol. 12, no. 10, 2020.

- [22] V. Venkatesh, J. Y. L. Thong, and X. Xu, "Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology," *MIS Q. Manag. Inf. Syst.*, vol. 36, no. 1, 2012, doi: 10.2307/41410412.
- [23] S. Olsson, J. Frellsen, W. Boomsma, K. V. Mardia, and T. Hamelryck, "Inference of structure ensembles of flexible biomolecules from sparse, averaged data," *PLoS One*, vol. 8, no. 11, 2013, doi: 10.1371/journal.pone.0079439.
- [24] P. W. Corrigan and J. R. O'Shaughnessy, "Changing mental illness stigma as it exists in the real world," *Aust. Psychol.*, vol. 42, no. 2, 2007, doi: 10.1080/00050060701280573.
- [25] J. E. Petit et al., "Response of atmospheric composition to COVID-19 lockdown measures during spring in the Paris region (France)," *Atmos. Chem. Phys.*, vol. 21, no. 22, 2021, doi: 10.5194/acp-21-17167-2021.
- [26] J. Scholz and K. Duffy, "We ARE at home: How augmented reality reshapes mobile marketing and consumer-brand relationships," *J. Retail. Consum. Serv.*, vol. 44, 2018, doi: 10.1016/j.jretconser.2018.05.004.
- [27] P. Ifinedo, "An empirical analysis of factors influencing internet/e-business technologies adoption by smes in Canada," *Int. J. Inf. Technol. Decis. Mak.*, vol. 10, no. 4, 2011, doi: 10.1142/S0219622011004543.
- [28] B. M. de Silva, D. M. Higdon, S. L. Brunton, and J. N. Kutz, "Discovery of Physics From Data: Universal Laws and Discrepancies," *Front. Artif. Intell.*, vol. 3, 2020, doi: 10.3389/frai.2020.00025.
- [29] A. Javornik, B. Marder, M. Pizzetti, and L. Warlop, "Augmented self - The effects of virtual face augmentation on consumers' self-concept," *J. Bus. Res.*, vol. 130, 2021, doi: 10.1016/j.jbusres.2021.03.026.
- [30] H. M. Beheshti and C. M. Beheshti, "Improving productivity and firm performance with enterprise resource planning," *Enterp. Inf. Syst.*, vol. 4, no. 4, 2010, doi: 10.1080/17517575.2010.511276.
- [31] R. B. Chase and S. Dasu, "Want to perfect your company's service? Use behavioral science," *Harv. Bus. Rev.*, vol. 79, no. 6, 2001.
- [32] R. Shamir, "Capitalism, governance, and authority: The case of corporate social responsibility," *Annu. Rev. Law Soc. Sci.*, vol. 6, 2010, doi: 10.1146/annurev-lawsocsci-102209-153000.
- [33] K. A. Arditte Hall, M. E. Quinn, W. M. Vanderlind, and J. Joermann, "Comparing cognitive styles in social anxiety and major depressive disorders: An examination of rumination, worry, and reappraisal," *Br. J. Clin. Psychol.*, vol. 58, no. 2, pp. 231-244, 2019, doi: 10.1111/bjc.12210.
- [34] J. Milward, P. Deluca, C. Drummond, R. Watson, J. Dunne, and A. Kimergård, "Usability testing of the BRANCH smartphone app designed to reduce harmful drinking in young adults," *JMIR mHealth uHealth*, vol. 5, no. 8, 2017, doi: 10.2196/mhealth.7836.
- [35] N. Urbach and B. Müller, "The Updated DeLone and McLean Model of Information Systems Success," 2012. doi: 10.1007/978-1-4419-6108-2\_1.
- [36] J. H. Wu and Y. M. Wang, "Measuring KMS success: A respecification of the DeLone and McLean's model," *Inf. Manag.*, vol. 43, no. 6, 2006, doi: 10.1016/j.im.2006.05.002.
- [37] V. McKinney, K. Yoon, and F. Zahedi, "The measurement of Web-customer satisfaction: An expectation and disconfirmation approach," *Inf. Syst. Res.*, vol. 13, no. 3, 2002, doi: 10.1287/isre.13.3.296.76.
- [38] A. Jeyaraj, "DeLone & McLean models of information system success: Critical meta-review and research directions," *Int. J. Inf. Manage.*, vol. 54, 2020, doi: 10.1016/j.ijinfomgt.2020.102139.
- [39] F. N. Machado-da-Silva, F. de S. Meirelles, D. Filenga, and M. B. Filho, "Student satisfaction process in virtual learning system: Considerations based in information and service quality from Brazil's experience," *Turkish Online J. Distance Educ.*, vol. 15, no. 3, 2014, doi: 10.17718/tojde.52605.
- [40] P. Yang et al., "Scattering and absorption property database for nonspherical ice particles in the near- Through far-infrared spectral region," *Appl. Opt.*, vol. 44, no. 26, 2005, doi: 10.1364/AO.44.005512.
- [41] E. Mazaheri, M. O. Richard, and M. Laroche, "Online consumer behavior: Comparing Canadian and Chinese website visitors," *J. Bus. Res.*, vol. 64, no. 9, 2011, doi: 10.1016/j.jbusres.2010.11.018.
- [42] Z. Jiang, J. Chan, B. C. Y. Tan, and W. S. Chua, "Effects of interactivity on website involvement and purchase intention," *J. Assoc. Inf. Syst.*, vol. 11, no. 1, 2010, doi: 10.17705/1jais.00218.
- [43] D. G. H. Divayana, "The implementation of blended learning with kelase platform in the learning of assessment and evaluation course," *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 17, 2019, doi: 10.3991/ijet.v14i17.8308.
- [44] I. Ariawan et al., "SARS CoV-2 Antibody Seroprevalence in Jakarta, Indonesia: March 2021," *SSRN Electron. J.*, 2021, doi: 10.2139/ssrn.3954041.
- [45] I. N. E. Mertayasa, I. G. B. Subawa, K. Agustini, and D. S. Wahyuni, "Impact of cognitive styles on students' psychomotoric abilities on multimedia course practicum," *J. Phys. Conf. Ser.*, vol. 1810, no. 1, pp. 1-9, 2021, doi: 10.1088/1742-6596/1810/1/012056.
- [46] A. C. Y. Chiang, E. McCartney, P. H. O'Farrell, and H. Ma, "A Genome-wide Screen Reveals that Reducing Mitochondrial DNA Polymerase Can Promote Elimination of Deleterious Mitochondrial Mutations," *Curr. Biol.*, vol. 29, no. 24, 2019, doi: 10.1016/j.cub.2019.10.060.
- [47] I. Muda and E. Ade Afrina, "Influence of human resources to the effect of system quality and information quality on the user satisfaction of accrual-based accounting system," *Contaduria y Adm.*, vol. 64, no. 2, 2019, doi: 10.22201/fca.24488410e.2019.1667.
- [48] R. H. E. MOUSTAFA, "THE ROLE OF EDUTAINMENT IN MUSEUMS, LEARN THROUGH PLAY," *Int. J. Multidiscip. Stud. Herit. Res.*, vol. 3, no. 1, 2020, doi: 10.21608/ijmsr.2020.180080.
- [49] G. Del Chiappa, L. Andreu, M. G. Gallarza, L.-L. Chang, K. F. Backman, and Y. C. Huang, "Emotions and visitors' satisfaction at a museum," *Int. J. Cult. Tour. Hosp. Res. Iss Christina Goulding Eur. J. Mark. Int. J. Cult. Tour. Hosp. Res. Iss Eur. J. Mark.*, vol. 88, no. 7, 2014.
- [50] J. Pallud, "Impact of interactive technologies on stimulating learning experiences in a museum," *Inf. Manag.*, vol. 54, no. 4, 2017, doi: 10.1016/j.im.2016.10.004.
- [51] T. Komarac, D. Ozretic-Dosen, and V. Skare, "Managing edutainment and perceived authenticity of museum visitor experience: insights from qualitative study," *Museum Manag. Curatorsh.*, vol. 35, no. 2, 2020, doi: 10.1080/09647775.2019.1630850.
- [52] S. Sandusky, "Gamification in Education," *Asbbs Am. Soc. Bus. Behav. Sci.*, vol. 21, no. 1, pp. 32-39, 2014, doi: 10.1007/978-3-319-10208-5.
- [53] S. Akter, J. D'Ambra, and P. Ray, "Development and validation of an instrument to measure user perceived service quality of mHealth," *Inf. Manag.*, vol. 50, no. 4, 2013, doi: 10.1016/j.im.2013.03.001.
- [54] D. vom Lehn and C. Heath, "Action at the exhibit face: video and the analysis of social interaction in museums and galleries," *J. Mark. Manag.*, vol. 32, no. 15-16, 2016, doi: 10.1080/0267257X.2016.1188846.
- [55] H. Swain, "Museum revolutions: how museums change and are changed," *Cult. Trends*, vol. 19, no. 3, 2010, doi: 10.1080/09548963.2010.495286.
- [56] M. Economou and L. P. Tost, "Evaluating the use of virtual reality and multimedia applications for presenting the past," in *Handbook of Research on Technologies and Cultural Heritage: Applications and Environments*, 2010. doi: 10.4018/978-1-60960-044-0.ch011.
- [57] T. Tchoubar, "Effective Use of Multimedia Explanations in Open e-Learning Environment Fosters Student Success," *Int. J. Inf. Educ. Technol.*, 2014, doi: 10.7763/ijet.2014.v4.370.
- [58] C. Pancioli, V. Russo, and A. Macaudo, "When Technology Meets Art: Museum Paths between Real and Virtual," 2017. doi: 10.3390/proceedings1090913.
- [59] P. D. Petrov and T. V. Atanasova, "The Effect of augmented reality on students' learning performance in stem education," *Inf.*, vol. 11, no. 4, 2020, doi: 10.3390/INFO11040209.
- [60] P. Balloffet, F. H. Courvoisier, and J. Lagier, "From museum to amusement park: The opportunities and risks of edutainment," *Int. J. Arts Manag.*, vol. 16, no. 2, 2014.
- [61] D. Wirtz, A. Tucker, C. Briggs, and A. M. Schoemann, "How and Why Social Media Affect Subjective Well-Being: Multi-Site Use and Social

- Comparison as Predictors of Change Across Time,” *J. Happiness Stud.*, vol. 22, no. 4, 2021, doi: 10.1007/s10902-020-00291-z.
- [62] E. J. M. Arruda-Filho, J. A. Cabusas, and N. Dholakia, “Social behavior and brand devotion among iPhone innovators,” *Int. J. Inf. Manage.*, vol. 30, no. 6, 2010, doi: 10.1016/j.ijinfomgt.2010.03.003.
- [63] C. M. Chiu, C. C. Chang, H. L. Cheng, and Y. H. Fang, “Determinants of customer repurchase intention in online shopping,” *Online Inf. Rev.*, vol. 33, no. 4, 2009, doi: 10.1108/14684520910985710.
- [64] W. S. Jeong et al., “Men with severe lower urinary tract symptoms are at increased risk of depression,” *Int. Neurourol. J.*, vol. 19, no. 4, 2015, doi: 10.5213/inj.2015.19.4.286.
- [65] D. Scarpi, “Work and Fun on the Internet: The Effects of Utilitarianism and Hedonism Online,” *J. Interact. Mark.*, vol. 26, no. 1, 2012, doi: 10.1016/j.intmar.2011.08.001.
- [66] T. Wagner and T. Rudolph, “Towards a hierarchical theory of shopping motivation,” *J. Retail. Consum. Serv.*, vol. 17, no. 5, 2010, doi: 10.1016/j.jretconser.2010.04.003.
- [67] G. Wang, I. S. Oh, S. H. Courtright, and A. E. Colbert, “Transformational leadership and performance across criteria and levels: A meta-analytic review of 25 years of research,” *Group and Organization Management*, vol. 36, no. 2, 2011. doi: 10.1177/1059601111401017.
- [68] K. Ryu, H. Han, and S. S. Jang, “Relationships among hedonic and utilitarian values, satisfaction and behavioral intentions in the fast-casual restaurant industry,” *Int. J. Contemp. Hosp. Manag.*, vol. 22, no. 3, 2010, doi: 10.1108/09596111011035981.
- [69] Y. S. Chen, “The positive effect of green intellectual capital on competitive advantages of firms,” *J. Bus. Ethics*, vol. 77, no. 3, 2008, doi: 10.1007/s10551-006-9349-1.
- [70] D. Z. Kuo, A. J. Houtrow, P. Arango, K. A. Kuhlthau, J. M. Simmons, and J. M. Neff, “Family-centered care: Current applications and future directions in pediatric health care,” *Maternal and Child Health Journal*, vol. 16, no. 2, 2012. doi: 10.1007/s10995-011-0751-7.
- [71] D. Gursoy, “Future of hospitality marketing and management research,” *Tour. Manag. Perspect.*, vol. 25, 2018, doi: 10.1016/j.tmp.2017.11.008.
- [72] L. V. Casaló, C. Flavián, and S. Ibáñez-Sánchez, “Influencers on Instagram: Antecedents and consequences of opinion leadership,” *J. Bus. Res.*, vol. 117, 2020, doi: 10.1016/j.jbusres.2018.07.005.
- [73] C. Tam and T. Oliveira, “Understanding mobile banking individual performance: The DeLone & McLean model and the moderating effects of individual culture,” *Internet Res.*, vol. 27, no. 3, 2017, doi: 10.1108/IntR-05-2016-0117.
- [74] P. A. Rauschnabel, “Virtually enhancing the real world with holograms: An exploration of expected gratifications of using augmented reality smart glasses,” *Psychol. Mark.*, vol. 35, no. 8, 2018, doi: 10.1002/mar.21106.
- [75] J. K. Hietanen, A. Myllyneva, T. M. Helminen, and P. Lyyra, “The effects of genuine eye contact on visuospatial and selective attention,” *J. Exp. Psychol. Gen.*, vol. 145, no. 9, 2016, doi: 10.1037/xge0000199.
- [76] C. Elston, “ICT in Primary Education,” *Using ICT Prim. Sch.*, vol. 3, no. September 2014, pp. 14–22, 2012, doi: 10.4135/9781446214343.n3.
- [77] A. H. Gold, A. Malhotra, and A. H. Segars, “Knowledge management: An organizational capabilities perspective,” *J. Manag. Inf. Syst.*, vol. 18, no. 1, 2001, doi: 10.1080/07421222.2001.11045669.
- [78] J. Henseler, C. M. Ringle, and M. Sarstedt, “A new criterion for assessing discriminant validity in variance-based structural equation modeling,” *J. Acad. Mark. Sci.*, vol. 43, no. 1, 2015, doi: 10.1007/s11747-014-0403-8.
- [79] M. Aparicio, F. Bacao, and T. Oliveira, “Grit in the path to e-learning success,” *Comput. Human Behav.*, vol. 66, 2017, doi: 10.1016/j.chb.2016.10.009.
- [80] H. Y. Lim, W. Wang, J. Chen, K. Ocorr, and R. Bodmer, “ROS regulate cardiac function via a distinct paracrine mechanism,” *Cell Rep.*, vol. 7, no. 1, 2014, doi: 10.1016/j.celrep.2014.02.029.
- [81] M. Y. C. Yim, S. C. Chu, and P. L. Sauer, “Is Augmented Reality Technology an Effective Tool for E-commerce? An Interactivity and Vividness Perspective,” *J. Interact. Mark.*, vol. 39, 2017, doi: 10.1016/j.intmar.2017.04.001.
- [82] K. Agustini, D. S. Wahyuni, I. N. E. Mertayasa, N. K. Wedhanti, and W. Sukrawarpala, “Student-centered learning models and learning outcomes: Meta-analysis and effect sizes on the students’ thesis,” in *Journal of Physics: Conference Series*, 2021, vol. 1810, no. 1. doi: 10.1088/1742-6596/1810/1/012049.
- [83] M. R. Habibi, M. Laroche, and M. O. Richard, “Brand communities based in social media: How unique are they? Evidence from two exemplary brand communities,” *Int. J. Inf. Manage.*, vol. 34, no. 2, 2014, doi: 10.1016/j.ijinfomgt.2013.11.010.
- [84] G. McLean and K. Osei-Frimpong, “Examining satisfaction with the experience during a live chat service encounter-implications for website providers,” *Comput. Human Behav.*, vol. 76, 2017, doi: 10.1016/j.chb.2017.08.005.
- [85] X. Guo, K. C. Ling, and M. Liu, “Evaluating factors influencing consumer satisfaction towards online shopping in China,” *Asian Soc. Sci.*, vol. 8, no. 13, 2012, doi: 10.5539/ass.v8n13p40.
- [86] K. Agustini, I. M. Putrama, D. S. Wahyuni, and I. N. E. Mertayasa, “Applying Gamification Technique and Virtual Reality for Prehistoric Learning toward the Metaverse,” *Int. J. Inf. Educ. Technol.*, 2022.
- [87] Y. Kim, Y. Park, and J. Choi, “A study on the adoption of IoT smart home service: using Value-based Adoption Model,” *Total Qual. Manag. Bus. Excell.*, vol. 28, no. 9–10, 2017, doi: 10.1080/14783363.2017.1310708.
- [88] K. Agustini, I. W. Santyasa, and I. M. Tegeh, “Quantum Flipped Learning and Students’ Cognitive Engagement in Achieving Their Critical and Creative Thinking in Learning,” *Int. J. Emerg. Technol. Learn.*, vol. 17, no. 18, pp. 4–25, 2022, doi: 10.3991/ijet.v17i18.32101.

# Developing a Computer Simulation to Study the Behavior of Factors Affecting the Flooding of the Gash River

Abdalilah G. I. Alhalangy

Department of Computer Sciences-College of Science and Arts-Ar Rass, Qassim University, Saudi Arabia

**Abstract**—In recent years, the city of Kassala has suffered from frequent flooding disasters in the Gash River, which is the city's lifeblood. But the problem of frequent flooding of the river has made it a life-threatening nightmare. The importance of research lies in the fact that it is one of the few attempts to discuss and study the causes and effects of the Gash River floods. It aims to identify the factors affecting river floods. It proposes an algorithm to simulate flooding by randomly generating different factors that effectively affect river flooding. The descriptive analytical approach, the analytical, inductive approach, and the analytical deductive approach to desk research were used, taking advantage of the primary statistical method in its observation and evaluation, which relies on primary and secondary information to help make scientific, practical, and objective. The research came out with significant results related to the problems that threaten the town of Kassala from the frequent floods of the Gash River. The study's results proved that there is a deviation and discrepancy between the floods rate during the year, which gives a negative indication, and that deposited quantities vary in different proportions from one period to another, which causes a significant threat in the future. The research suggests other solutions that help reduce the problems and their effects. In addition to the above, the study proposes various recommendations that will be the basis for future studies to reach the required solutions and goals.

**Keywords**—Gash river; flood simulation; influencing factors to flood; simulation

## I. INTRODUCTION

Floods frequently result in significant economic losses and harm to people and the community [1]. Flood forecasting and management have thus always been difficult for the government and local governments [2, 3]. As more and more computer models are used for flood management [4], there is an increasing need for innovative and effective ways to visualize large amounts of data about floods. In addition, because interaction is one of the most critical factors in modern decision-making tools, simulation effectiveness has become a significant factor [5, 6]. Flooding in urban contexts is typically associated with a complex landscape involving technical, social, economic, and environmental issues [7]. The Gash River is a seasonal river that descends from the Eritrean plateau and reaches the Sudanese border, 30 km south of Kassala. It passes through the city of Kassala, passing through the Gash Delta Agricultural Project, to end at the Gash Delta, about 91 km north of the area of Kassala [8]. The natural features of the Gash River are unique characteristics,

including the steep slope, which is six times the sharpest stream in Sudan, as its slope within the Sudanese borders is 1.3 m / km at Kassala city, which gave it the ability to bring high loads of siltation [9].

Soft, sandy soils and fragile banks characterize the riverbed banks. It is easily eroded by high currents and fragmentation and change course from one place to another. These factors combined lead to the occurrence of flood phenomenon [7]. The area of the basin of the Gash River is estimated at 21.000 km<sup>2</sup>. Most of it is located within the Eritrean lands and at contour heights ranging between 2000 and 1100 m above sea level. This basin is located between longitudes 36.5 and 39.5 east and latitudes 14 and 15.5 north. The river in Eritrean territory has a steep slope that reaches 5.5 meters per kilometer. The transition from about 2000 meters above sea level at its source to only about 550 meters at the Sudanese border [9] makes the flood waves rush very quickly. It reaches 2 meters/sec sometimes [8]. Therefore, the Gash is among the fastest rivers in the world [10]. The river's slope within the flat Sudanese plain is estimated at 1.3 m / km.

The importance of the study lies in the fact that it reveals the role of technical means in reducing water losses due to Flooding and benefiting from them in drinking, agriculture, and creating new projects in the agricultural field. And this study takes a fresh look at the country's flood management, and explores methods for lessening the likelihood of flooding and the damage it might cause [11].

In addition to reflecting technical progress, designing a computer simulation model for the behavior of factors affects flooding.

The Gash River has become a source of fear and anxiety because it floods and damages the economy, society, and environment. These floods are caused by many factors work together. Which is difficult to predict and control, and whatever the matter is, floods are a natural phenomenon that spreads in all parts of the world, and this is not the dilemma, but rather the difficulty remains in that we avoid its occurrence and prevent the disaster before it occurs and to what extent we contribute to it [12].

The study is concentrated in Kassala State. Kassala State, its capital, is the city of Kassala. The area of Kassala State is 55.374 square kilometers, and it is located between latitudes 12-34 and 36-57 east and longitudes 12-14 and 12-17 north. In the east, the state's borders are shared with the country of

Eritrea. Internally, it borders the Red Sea, Gezira, and Gedaref states [13]. The state enjoys a prominent geographical location that helped its growth and made it a center for many commercial and investment activities. It provides essential services and all factors that help investment. It is known for its beautiful scenery and lush gardens, which make it the most important place for tourists to visit. There are many agricultural projects, such as sugar cane production, corn and wheat too. The populace also engages in trading and pastoral work.

## II. MORPHOLOGY OF THE GASH RIVER

The Gash River is known from the morphological point of view as a cascading river, where the river is wide and shallow. In this type of river, the current takes several directions, is unstable, and constantly changes its course and characteristics [8].

The Gash River carries large quantities of Silt (primarily fine sand and Silt). Due to the sharp decrease in the river's slope when it crosses the Sudanese lands, these quantities exceed the river's ability to transport it. It is deposited along the course of the river until its end, which caused the river's steadily reducing capacity and inability to transport it. Passing High Floods Safely [9].

## III. METHODOLOGY

The research method used in the study is the descriptive analytical method, and this approach is also called the in-depth descriptive method. In this method, the research describes various scientific phenomena and problems in detail, allowing good explanations and results. Using the descriptive-analytic approach, different things are compared to similar items. This allows data to be collected about the differences and similarities between things, which are the most important things that make the descriptive-analytical approach different from other scientific approaches [15].

The descriptive-relational approach is also used because it is related to the nature of the study. It also relies on analyzing the existing system and using diagrams, graphs, flowcharts, algorithms, design, and software to create a more accurate and efficient system. According to the explanatory method, relational studies define problems and analyze them into their essential elements [15] by presenting events and facts and analyzing them. Also, the inductive-analytical approach involves immersing in the details and reasoning behind the points to identify significant patterns, themes, and interrelationships, after which the results are confirmed using analytical principles rather than rules [14]. The study also follows the deductive analytical desk research approach, in which the literature related to the subject of the study is reviewed.

## IV. MODEL DESCRIPTION AND ALGORITHM

To study the current situation, the study dealt with the problem of the Flooding of the Gash River by simulating the factors affecting the Flooding in the form of randomly generated variables within the program environment.

### A. Factors Affecting the Flooding of the Gash River

There are many factors affecting the Flooding of the Gash River. Still, they can be limited to natural, technical, and other artificial causes attributed to some harmful practices.

### B. Flood Estimate

The Gash River is distinguished for its wide fluctuations in its water revenue throughout and between seasons (Table I). Evidence indicates changes in the river's hydrology as a result of changing the natural properties of the upper river basin in the State of Eritrea, especially the removal of vegetation and soil erosion, which made the flood wave more severe and shorter in time. This is in addition to the global climatic changes that began heralding the high precipitation rates in 2007 AD [9].

TABLE I. SHOWS THE RIVER'S DISCHARGES FROM 1997-2010 (AL-JIRA STATION) [8]

Year	Total Revenues Upstream - Million Cubic Meters	Year	Total Revenues Upstream - Million Cubic Meters
1997	111.58	2009	452.57
1998	565.21	2010	680.69
1999	1386.66	2011	1007.41
2000	1240.00	2012	704.04
2001	1007.41	2013	888.89
2002	704.04	2014	1313.09
2003	1449.49	2015	1300.66
2004	874.97	2016	440.00
2005	1101.00	2017	760.00
2006	1207.75	2018	663.80
2007	1760.00	2019	450.85
2008	665.20	2020	673.02

We note from the above table the great variation in the river's revenues during the season. We also notice the difference in the discharges and returns of the river during a number of selected years at the Al-Jira monitoring station at the source (Fig. 1) illustrates this.

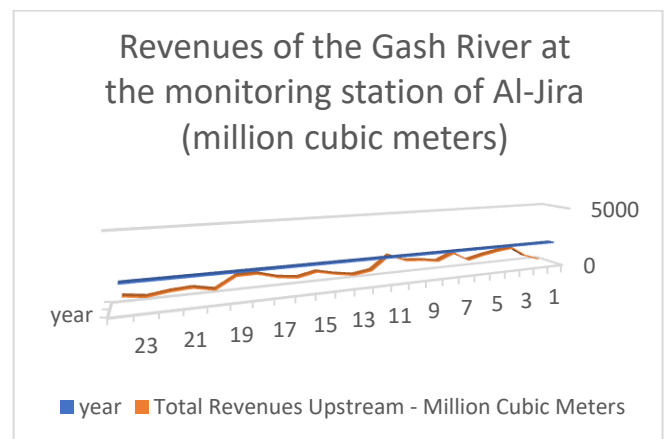


Fig. 1. Shows the discharges and revenues of the river at the Jira station upstream.

Flood estimation methods aim at modeling the relationship between falling rain and the resulting surface runoff. In general, these methods can be divided into three groups. The first group is called simple methods - such as the logical method and implicit equations - from which the value of the peak behavior can be estimated quickly and based on a Small amount of information. The CN Curve Number method is one of the methods of the second group or group of medium complexity [16]. The third group includes straightforward or more complex methods, which depend on the study of each element involved in the phenomenon, the transformation of rain into a surface runoff, and detailed and in-depth analysis. The number curve method is the most prevalent in engineering studies and natural resource management projects, especially in the US states [17].

C. *The Nature of the Gash River (The Strength of the Gash River Current)*

The significant fluctuation represents the characteristics of the Gash River in the flow, the steepness, the high speeds, high loads of Silt, the broad stream, and the fragile banks combine to form a community that threatens the river's bridges and its widths.

D. *Silt*

Problem of Silt is one of the most important problems associated with irrigation, drainage, floods, and natural rivers. Generally, dredged Silt represents about 5 to 25% of the total Silt carried, but its importance is that it affects the roughness and stability of the stream. Table II shows the quantities of dredged Silt during periods of years from 2010 to 2020.

As can be seen in Table III, the amounts of Silt vary significantly not only from season to season, but even within a season.

TABLE II. SHOWS THE QUANTITIES OF DREDGED SILT DURING THE PERIOD FROM 2010 TO 2020

Year	Total - mm3	The upper limit is 3 mm
2010	1290	470
2011	0930	385
2012	0810	370
2013	1430	870
2014	0400	205
2015	0535	365
2016	1165	575
2017	1010	395
2018	0960	290
2019	0540	295
2020	0540	295

TABLE III. SHOWS THE DIFFERENT AMOUNTS OF SILT

Date	Attributed	Suspended Silt (ppm)
21/08/2020	508.0	1753.7
22/08/2020	507.0	1450.5
25/08/2020	506.9	13458
27/08/2020	506.8	13300
29/08/2020	506.6	11976

The table shows that silt quantities vary seasonally and within seasons.

The amount of Silt in the water decreases according to the amount of water (water levels). The higher the group, the higher the amount of Silt attached to the water. We find the rise of the stream bed in the two divided sectors of the river, in front of and behind the network of bridges and wide's, receiving their share of the Silt carried from the upper reaches of the river [9].

E. *Various other Reasons*

There are many harmful practices carried out by the population that affects negatively, including some activities around the river basin have adverse effects such as planting trees around bridges, erecting buildings, constructing bridges and beams inside the bay, taking soil from bridges, or digging latrines and wells in the areas adjacent to the bridges [18].

F. *Elevation of the Levels of the Gash River*

This is due to three main reasons:

First: the climatic and hydrological changes represented in the disappearance of the vegetation cover and the increase in water and silt discharges, the engineering works that prevented the natural venting that was occurring in the Tajuj region at high levels, and the collapse of the dams built on the Gash River in the State of Eritrea.

Second: Weakness of the conveying capacity of the gutter as a result of the presence of suffocating facilities such as bridges and bridges.

Third: the morphological changes that led to the rise of the riverbed due to sedimentation due to the lack of slope.

All these reasons combine to cause a catastrophe, as happened in 2003. Climate change is a global phenomenon that warns of more variation in precipitation and high waves of floods. The rise in the river water level as result of heavy rains was accompanied by the collapse of some dams in the country of Eritrea, which exacerbated the severity of the floods [8].

---

Program algorithm generating random numbers to simulate the factors influencing Flooding

---

```

1. Start
2.1 let I= 1
2.2 sum=0
2.3 pita=300
3. Let      Z=random (10)
4. If (z = 0)
{x=0}
Else
{x=pita*Z}
5. Near the value x to an integer
6. If (x=0)
{M=pita/4 Go to 8}
7. Let M=ceil(x)*1/4
8. Sum = sum + ceil(x)
9. Print value ceil (m), ceil(x)
10. Let I=I+1
11. If (I<=10)
{Go to 3}
12. Print Report
13. End
    
```

---



In the algorithm

I: represents the loop counter.

Pita: means the amount of water.

Z: represents the randomly generated number from the random function. The random process generates random numbers for unrelated events; the Ceil function approximates the correct number.

M: represents the amount of Silt, which is calculated by the formula  $m = \text{pita}/4$ .

Fig. 2 shows the sequence of steps of the above program and algorithm through the following flowchart.

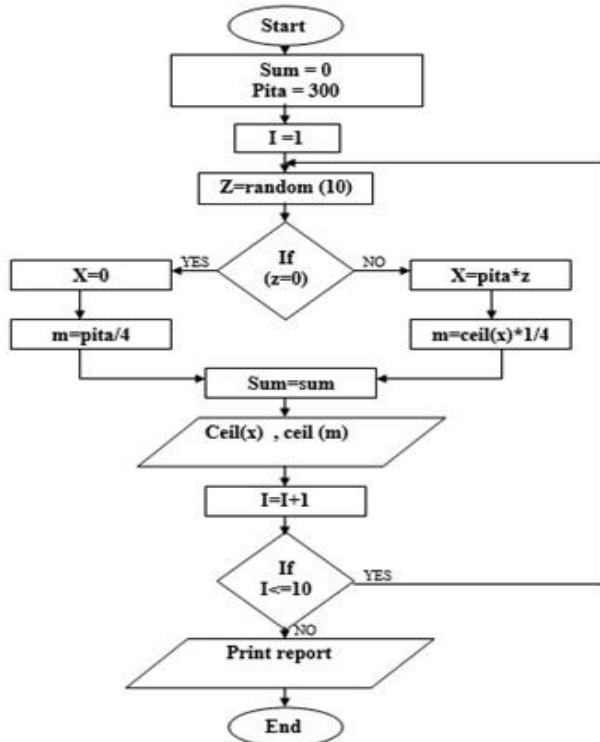


Fig. 2. Flow chart of algorithm.

Fig. 3 illustrates the implementation of the program, and we find that the sediment quantities vary in different proportions from one period to another, which causes a great threat in future periods.

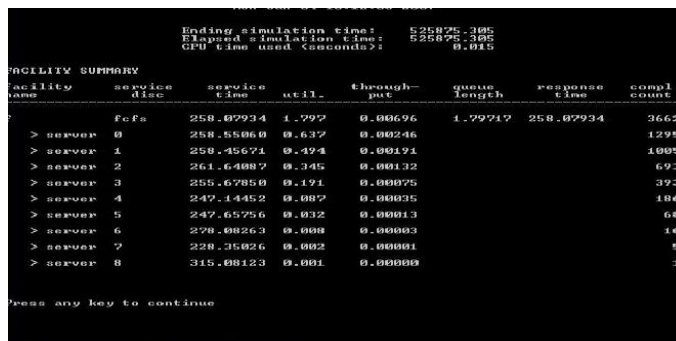


Fig. 3. Shows the program execution screen.

TABLE IV. SHOWS THE OUTPUT OF THE PROGRAM

Year	M (The amount of clay)	X (Quantity of Water)
1	1800	450
2	2400	600
3	0300	075
4	1800	450
5	1200	300
6	2400	600
7	0600	150
8	1800	450
9	0300	075
10	1800	450

Table IV shows a reflection of the implementation of the program shown in (Fig. 3) and shows the relationship between the amount of water in the river and the amounts of suspended silt for ten consecutive years generated randomly.

The total amount of Silt after 10 years is 3600 m.

## V. IMPLEMENTATION

A Markov distribution was used to generate random numbers and compare unrelated cases with equal probability distribution using the beta parameter calculated from the entire population.

A sample of 10 years was taken to take the average from the actual reality of the cascade generation that occurs annually starting from the base year. Table V shows the average level of the river level during specific years at the Al-Qash Gash Bridge monitoring station in the city center, as well as the lowest rate, highest rate and the standard deviation.

By comparing the result of the program and the actual reality, we find the relationship between the level and the Silt is correct.

Benefit from the program in calculating the amount of Silt deposited after each number of years.

TABLE V. AVERAGE RIVER LEVEL DURING SELECTED YEARS - AL-GASH BRIDGE MEASURING STATION

Year	The Lowest Rate	Highest Rate	Standards deviation
1990	15.00	210.00	31.2842
1999	504.00	506.00	0.3167
2003	504.00	505.90	0.2563
2005	504.05	505.95	0.3553
2008	505.40	506.01	0.2986
2010	504.70	505.95	0.1951
2015	504.90	506.10	0.2346
2019	504.19	506.45	0.3481

## REFERENCES

The results of the study proved that there is a large deviation and variation between sedimentation rates during the year, which gives a negative indicator, and that the amounts of sediment vary in varying proportions from one period to another, which causes a great threat in future. It is also expected that the amount of Silt deposited by this algorithm will be about 8.450 mm<sup>3</sup> in the year 2035 AD. Different river speeds, revenues, and flows reveal a clear relationship between the volume of water and its rate and the amount of Silt drifting and traveling from one place to another within the same season. The study also proved that the annual average of rain is not considered a good factor for the strength of the rainstorm in estimating the severity of the flood, as it is possible for the power of one of the rainstorms to exceed this annual average.

The study also concluded that increased sediment quantities over the years led to a rise in the river level higher than the level of agricultural lands. This was one of the primary reasons and systematic risks that allowed the river to stray from its intended path. Changing the course of the river may pose a danger in the future.

## VI. CONCLUSION

Developing this model includes all elements have an indirect effect, such as evaporation, soil type, and wind.

The model works on developing a distant understanding, a broad horizon, and an open mind in the applied aspects at all administrative levels, especially about the applications of simulation models.

Also, this model benefits in the field of forecasting and future estimates of the quantities of silt and water for the Gash River.

The application of this model in the investigation and exploration of the river is a powerful technique through which we can provide planners of water resources with an acceptable estimate of the amount of running water and the quantities of suspended Silt that are expected to be deposited.

The results of the model can also be used to help city planners and decision-makers make plans for the future of Kassala city; applying this modern technical method (simulation models) in flood management in the country.

The researcher suggests that there will be future studies that follow the technical method on the basin's area, the mainstream's length, and the peak discharge.

It is also suggested that there be studies of the geological formation, soil permeability, and type of land use around the course of the Gash River to avoid the risks of floods in the future.

Similarly, more in-depth studies reach more detailed results to advance this region and develop it in the future in terms of economic and social aspects.

- [1] Muhammad Hafizi Mohd Ali, Siti Azirah Asmai, Z. Zainal Abidin, Zuraida Abal Abas and Nurul A. Emran, "Flood Prediction using Deep Learning Models" International Journal of Advanced Computer Science and Applications(IJACSA),13(9),2022. <http://dx.doi.org/10.14569/IJACSA.2022.01309112>
- [2] G. V. Merkurjeva and M. Kornevs, "Water Flow Forecasting and River Simulation for Flood Risk Analysis," Information Technology and Management Science, vol. 16, no. 1, pp. 42-46, 2013.
- [3] P.-K. Chiang, P. Willems, and J. Berlamont, "A conceptual river model to support real-time flood control (Demer River, Belgium)," River Flow 2010, pp. 1407-1414, 2010.
- [4] Aslinda Hassan, Haniza Nahar, Wahidah Md Shah, Azlianor Abd-Aziz, Sarah Afiqah Sahiran, Nazrulazhar Bahaman, Mohd Riduan Ahmad, Isredza Rahmi A. Hamid and Muhammad Abu Bakar Sidik, "Performance Evaluation of Raspberry Pi as an IoT Edge Signal Processing Device for a Real-time Flash Flood Forecasting System" International Journal of Advanced Computer Science and Applications(IJACSA),13(10),2022. <http://dx.doi.org/10.14569/IJACSA.2022.01310100>
- [5] D. Cornel et al., "Interactive visualization of flood and heavy rain simulations," in Computer Graphics Forum, 2019, vol. 38, no. 3, pp. 25-39: Wiley Online Library.
- [6] J. Waser et al., "Many plans: Multidimensional ensembles for visual decision support in flood management," in Computer Graphics Forum, 2014, vol. 33, no. 3, pp. 281-290: Wiley Online Library.
- [7] M. G. MIGUEZ, F. C. B. MASCARENHAS, and A. P. VERÓL, "MODCEL: a mathematical model for urban flood simulation and integrated flood control design," in Proceedings IV Conference "Acqua e Città", Centro Studi Idraulica Urbana, Venice (Italy), 2011.
- [8] Ministry of Irrigation and Water Resources - Groundwater and Valleys Department, Kassala State. "Technical report of the Gash River Basin in the light of monitoring and follow-up data," 2020.
- [9] Higher Technical Committee for Taming the Gash River and Protecting the City of Kassala. Internally published reports. Kassala State "Technical report of the Gash River," 2021.
- [10] M. M. Khogli, "Strategies and efforts to address the impact of the Gash River floods on the city of Kassala," Bulletin of Sudanese Studies, vol. 16, no. 1, 2010.
- [11] Jain, Sharad K., and Vijay P. Singh. "Strategies for flood risk reduction in India." *ISH Journal of Hydraulic Engineering* (2022): 1-10.
- [12] Yusoff, Sarina, and Nur Hafizah Yusoff. "Disaster risks management through adaptive actions from human-based perspective: case study of 2014 flood disaster." *Sustainability* 14.12 (2022): 7405.
- [13] UNDP report, Kassala State, unpublished reports, "Situational analysis . Kassala City " 2021.
- [14] Muhammad Tayseer, "The Analytical Descriptive Method Book: With an Introduction to the Analytical Descriptive Method," in the Arab Journal for Science and Research Publishing Foundation, 2022. <https://blog.ajsrp.com/?p=35302>; <https://blog.ajsrp.com/?p=5663>
- [15] A.-F. A. Q. A. A.-S. Al-Fadni, "Scientific research method," 3rd edition Khartoum, p. p. 62., 2004.
- [16] G. M. Dawod, M. N. Mirza, and K. A. Al-Ghamdi, "GIS-based estimation of flood hazard impacts on the road network in Makkah city, Saudi Arabia," *Environmental Earth Sciences*, vol. 67, no. 8, pp. 2205-2215, 2012.
- [17] G. M. Dawod, M. N. Mirza, and K. A. Al-Ghamdi, "GIS-based spatial mapping of flash flood hazard in Makkah City, Saudi Arabia," *Journal of Geographic Information System*, vol. 3, no. 03, p. 225, 2011.
- [18] N. I. Adam, "The role of official and voluntary bodies in disaster prevention and reconstruction," 2019.

# Pinpointing Factors in the Success of Integrated Information System Toward Open Government Data Initiative: A Perspective from Employees

Wahyu Setiawan Wibowo<sup>1</sup>, Ahmad Fadhil<sup>2</sup>, Dana Indra Sensuse<sup>3</sup>, Sofian Lusa<sup>4</sup>,  
Prasetyo Adi Wibowo Putro<sup>5</sup>, Alivia Yulfitri<sup>6</sup>

Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia<sup>1,2,3,4,5</sup>

Faculty of Computer Science, Universitas Esa Unggul, Jakarta, Indonesia<sup>6</sup>

**Abstract**—As the Supervisory Institution in Statistics, Badan Pusat Statistik (BPS) launched an integrated information system (IS) to exercise the Open Government Data (OGD) initiative and to impose the One Data Policy Act. Albeit challenges arise, BPS manages to provide more than 120 thousand publicly accessible datasets. With the success of OGD, many scholars have opted to examine a similar issue from the perspective of users/citizens. However, employees' perspective remains substantial as employees are the OGD provider. This research administers employees' views to pinpoint influencing factors in the success of OGD adoption through an IS. The authors seek to comprehend the factors from IS and acceptance manner, thus integrating the Information System Success Model (ISSM) and Unified Theory of Acceptance and Use of Technology (UTAUT) as the measurement model. This study also administers a cross-sectional questionnaire with close-ended questions to obtain data from 253 IS users in BPS. Using structural equation modelling (SEM), the authors find that all ISSM constructs influence the success of IS while only one construct from UTAUT plays a pivotal role in defining the success. Information Quality, System Quality, Service Quality, User Satisfaction, and System Use remain paramount to the successful implementation, while Performance Expectancy becomes the sole influencing UTAUT factor affecting success. This study therefore offers substantial benefits by aiding other researchers in OGD-related areas and providing in-depth evidence for practitioners in implementing IS for OGD initiatives.

**Keywords**—Open data; open government data; OGD; employees' perspective; ISSM; UTAUT; success factors; acceptance; impact; integrated IS; One Data Policy; BPS; SEM

## I. INTRODUCTION

Since its inception in 2009 [1], [2], Open Government Data (OGD) have grown exponentially [3] in the last few years and has served as the bedrock for a data-driven nation. Correspondingly, governments have collected a plethora of data to perform their tasks and made it available to the public [4]. OGD therefore holds an increasingly pivotal role [5]. In 2016, President Joko Widodo coined One Data Initiative or "Satu Data Indonesia" (SDI), followed by a regulation in 2019: Presidential Decree Number 39 of 2019. Similar to principles of open data by the Open Knowledge Foundation [1] and the United Kingdom's Government [6], Indonesian ministries/agencies must comply with four fundamental principles (*data standard, metadata, interoperability, and*

*reference code*). Through SDI, the government intends to deliver accurate, up-to-date, integrated, accountable, accessible, and interchangeable data for national development agendas [7].

As SDI's Central Data Supervisory Institution (Pembina Data) in Statistics, Badan Pusat Statistik (BPS) introduced<sup>1</sup> an integrated information system (SIMDASI) that administers SDI principles: interoperability with other ministries/agencies [8]–[10] and compliance to data standards, metadata, and reference code<sup>2</sup>. Supported by derivative policies, SIMDASI establishes itself as the bedrock of the OGD initiative, dispensing most of the open data on the organisation's website<sup>3</sup> with the finest quality through synchronisation processes [8]. Presently, SIMDASI houses more than 120 thousand publicly accessible datasets<sup>4</sup>; thus, SIMDASI remains paramount in the OGD initiative for BPS.

Likewise, Information and Communications Technology (ICT) adoption for employees becomes tedious [11] and information systems (IS) cannot engender benefits should employees refuse or fail to use IS accordingly [12]. In measuring the success of government systems or services, Martono et al. (2020) [13], Puspitarini & Ardhani (2022) [14], Stefanovic et al. (2016) [15], and Gangga Dewi & Fajar (2021)[16] successfully exercised DeLone and McLean Information System Success Model (ISSM) [17] to measure the IS success. Accordingly, Stefanovic et al. (2021) [18] reaffirm that ISSM is the most widely used measurement to measure IS success within the e-government context. Stefanovic et al. (2016) [15] also evaluate the IS success of an e-government system from the view of employees in the Republic of Serbia. The result shows that the use of an e-government system influences the net benefits perceived by the employees.

At the organisational level, moreover, many scholars stress the challenges hindering OGD adoption: lack of technical skills

<sup>1</sup> <https://nasional.kontan.co.id/news/bps-luncurkan-sistem-data-statistik-terintegrasi>. Accessed: 7 November 2022

<sup>2</sup> <https://www.bps.go.id/menu/8/Peraturan.html>. Accessed: 7 November 2022

<sup>3</sup> [https://www.bps.go.id/indikator/indikator/list/\\_da\\_01/](https://www.bps.go.id/indikator/indikator/list/_da_01/). Accessed: 7 November 2022

<sup>4</sup> <https://simdasi.bps.go.id/>. Accessed: 7 November 2022

from employees, staff shortage, and organisational understanding [19]–[25]. Wang et al. (2019) [24] even address that the United Kingdom's OGD programme relies on employees' enthusiasm, skills, and goodwill. Subedi et al. (2022) [26] also imply that individual and organisational factors influenced OGD adoption by Nepali government officials. Hence, the acceptance from employees has proven substantial in the OGD adoption.

In assessing the acceptance, the Unified Theory of Acceptance and Use of Technology (UTAUT), developed by Venkatesh et al. (2003) [27] and perfected by Venkatesh et al. (2012) [28], becomes one of the baseline models to explain user intention to use an IS in an organisation setting such as in e-government research [12], [26], [29]–[38]. Their findings prove that the four construct (Performance Expectancy, Effort Expectancy, Facilitating Conditions, and Social Influence) affect the intention to use, which subsequently influences the system use. Further, Stefanovic et al. (2021) [18] explain that many scholars combine UTAUT in the ISSM construct to complement their studies [12], [14], [29], [38], [39].

In OGD-related research, however, many scholars employed the citizens' perspective [31], [37], [39]–[43] to measure the success of OGD, although Syuhaini and László (2022) [12] highlight the importance of taking employees' perspective as they design, implement, run, and expand the systems. Talukder et al. (2019) [43] propose additional dimensions from ISSM to clarify the usage intention of OGD users. Prior studies have yet to discuss the benefits of IS as the foundation of OGD and employees' acceptance as OGD providers for the public; they stop at the behavioural intention to use OGD. Considering the gaps and the limited number of research on OGD from employees' perspective, this study formulates the following research questions:

RQ1. What are the IS factors that influence the use of SIMDASI?

RQ2. What acceptance factors from employees affect the use of SIMDASI?

RQ3. What direct factors determine the success of SIMDASI as an OGD adoption tool?

To address the research question, the authors propose an UTAUT-integrated ISSM model aimed at unearthing benefits from administering SIMDASI as an OGD adoption tool from employees' perspective. The model

The study therefore is organised as follows: Section II elaborates theoretical background and hypotheses; Section III explains the research methodology; Section IV expounds on the research results; Section V explicates the discussion regarding the research objectives and the implications; Section VI closes the research with the conclusion and limitation of the study.

## II. LITERATURE REVIEW AND HYPOTHESES

### A. DeLone & McLean Information System Success Model (ISSM)

ISSM offers a comprehensive metric to gauge the success of an information system [44], [45]. Correspondingly, this approach proposes an idea in which system quality and information quality affect the use and the adoption of IS, resulting in individual and organisational impact [43], [45]. DeLone and McLean (2003) [17] further propounds upgrades with the inclusion of the service quality dimension to help explain the effectiveness of an IS [45]. DeLone and McLean (2016) [46] then recommended the integration of *intention to use* and *use* into the *system use* dimension, of which the model measures the degree of usage in a mandatory IS —*the case of SIMDASI*. The ISSM model thus stipulates six dimensions as follows:

#### 1) Information Quality (IQ)

Providing users with accurate, timely, and relevant information becomes the driving force for IS applications; hence, the IQ dimension remains paramount in the model [45]. DeLone and McLean (2016) [46] argue that IQ contributes to user satisfaction and declares that this dimension should be viewed as a success metric independent from end-user satisfaction measurements. Further, they assert that IQ has been demonstrated to be substantially correlated with system usage and net benefits in both individual and organisational settings [44]–[46].

Stefanovic et al. (2016) [15] and Martono et al. (2020) [13] also prove that IQ influences the usage of e-government systems. Correspondingly, Talukder et al. (2019) [43] pinpoint that IQ influences citizens' usage of OGD. For the satisfaction issue, Gangga Dewi & Fajar (2021) [16] verify that IQ affects employees' satisfaction in using IS at Kominfo, Indonesia, similar to the research by Puspitarini & Ardhani (2022) [14].

This study therefore refers to the IQ dimension as a measurement of SIMDASI following the criteria: *relevance, usefulness, understandability, accuracy, reliability, completeness, and timeliness* [45], [46]. Accordingly, the authors suspect that IQ influences system use and user satisfaction.

#### 2) System Quality (SQ)

DeLone and McLean (2003) describe that SQ constitutes the desired attributes of an IS; thus, SQ subsumes measures of the IS itself [44]–[46]. These measures typically emphasise usability aspects and performance attributes of the examined IS. DeLone and McLean also reiterate that there is no universal measurement in SQ; thus, they establish *ease of learning, ease of use, availability, system reliability, and system interactivity* as ways to gauge the SQ [44]–[46].

In this model, SQ influences the system use and user satisfaction of the IS examined. Puspitarini & Ardhani (2022) [14], Talukder et al. (2019) [43], and Stefanovic et al. (2016) [15] demonstrate how SQ affects both *intention to use* and *user satisfaction* in their research. In addition, Martono et al. (2020) [13] uncovered that SQ has a significant influence on *system usage*, and Gangga Dewi & Fajar (2021) [16] have verified the correlation between SQ and user satisfaction in their research.

The authors suspect SQ significantly influences *user satisfaction* and *system use* in the model.

### 3) Service Quality (SEQ)

SEQ, the updated ISSM dimension, denotes the quality of assistance provided to users by the IT support team or IS department, such as training, hotline, or helpdesk [44]–[46]. DeLone and McLean (2016) [46] reclarify the SEQ dimension, which they believe is misunderstood and understudied, merits inclusion as a component of IS success because SEQ enhances individual performance. They further propose that the key measures for today's digital environment are *reliability*, *empathy*, *responsiveness*, *contact*, and *interactivity*.

Puspitarini & Ardhani (2022) [14] and Gangga Dewi & Fajar (2021) [16] evince that SEQ influences both system use and user satisfaction. Stefanovic et al. (2016) [15], on the other hand, verify the relationship between SEQ and system usage in their research but fail to identify the correlation between SEQ and user satisfaction. The same goes for Martono et al. (2020) [13], as they failed to determine the connection between SEQ to either user satisfaction or system use.

As SIMDASI remains a mandatory system for OGD adoption and a specific department is available to assist users with technical and nontechnical support [8], the authors believe that the inclusion of SEQ is beneficial and thus argue that SEQ has a significant influence on system use and user satisfaction.

### 4) System Use (SU)

DeLone and McLean (2016) [46] explain that SU is the degree and manner in which employees utilise IS capabilities. Further, they also rectify the misconceptions among many scholars to omit SU where the system is mandatory. Realising the system's benefits is expected to be significantly impacted by variations in quality and intensity [46].

While prior studies stop at the SU, Stefanovic et al. (2016) [15] and Gangga Dewi & Fajar (2021) [16] seek to uncover the relationship between SU and the net benefits, which leads to the realisation that SU has a significant influence on the net benefits either for organisations or individuals.

The authors then suggest using SU as the mediating factor from IQ, SQ, and SEQ to the net benefits. Accordingly, DeLone and McLean (2016) [46] highlight indicators to measure SU: *frequency of use*, *duration of use*, *nature of use*, *appropriateness of use*, *number of functions or features used* (*extent of use*), *thoroughness of use*, *attitudes toward use*, and *intention to reuse*.

### 5) User Satisfaction (US)

The US dimension constitutes users' satisfaction when utilising an IS and is considered an essential factor in

measuring IS success. As SIMDASI remains mandatory, employing this dimension becomes favourable [44]–[46]. Gangga Dewi & Fajar (2021) [16] have proven the relationship between US to SU and US to net benefits, further proving the notion from Stefanovic et al. (2016) [15]. Urbach and Müller (2012) [47] also argue that US has a strong relationship with both system use and net benefits. The authors believe that US mediates IQ, SQ, and SEQ to SU and net benefits. The measurement of *effectiveness*, *enjoyment*, and *overall satisfaction* is an effective way to gauge US in this model. This study hence employs this US dimension as part of the measurement model.

### 6) Individual Impact (II)

DeLone and McLean (2016) [46] splits net impacts into *individual* and *organisational impacts*. This dimension explains the extent to which IS contributes to an individual's success, such as improved personal productivity. This paper suggests the inclusion of the II construct to measure the success of SIMDASI among individual levels that attribute the success of OGD adoption. The measurement of *learning*, *decision quality*, *decision time*, *productivity*, or *task performance* is proven effective.

### 7) Organisational Impact (OI)

DeLone and McLean (2016) [46] also propose measuring *cost reduction*, *overall productivity*, *improved outcomes*, and *e-government positioning* through the OI construct at the organisation level. As this study intends to unravel the success of SIMDASI to OGD initiatives in BPS, the authors propose using OI in the measurement model.

Given the above, the authors establish the proposed hypotheses to measure IS usage in the following lists:

H1. IQ positively and significantly influences on SU to operate SIMDASI for OGD adoption.

H2. IQ positively and significantly influences on US to operate SIMDASI for OGD adoption.

H3. SQ positively and significantly influences on SU to operate SIMDASI for OGD adoption.

H4. SQ positively and significantly influences on US to operate SIMDASI for OGD adoption.

H5. SEQ positively and significantly influences on SU to operate SIMDASI for OGD adoption.

H6. SEQ positively and significantly influences on US to operate SIMDASI for OGD adoption.

H7. US positively and significantly influences on SU to operate SIMDASI for OGD adoption.

### B. The Unified Theory of Acceptance and Use of Technology (UTAUT)

The UTAUT, a modified TAM model, uses two variables from TAM and expands them into four constructs (performance expectancy, effort expectancy, social influence, and facilitating conditions) affecting the behavioural intention to use a technology and/or technology use [27], [28]. Venkatesh et al. (2003) [27] imbue the construct with eight IT adoption theories, perfecting the theory. The eight standards

used in the construct are the Theory of Reasoned Action (TRA), Technology Acceptance Model (TAM), Motivational Model (MM), Theory of Planned Behaviour (TPB), a model that combined TAM and TPB, Model of PC Personal Computer Utilization (MPCU), Innovation Diffusion Theory (IDT), and Social Cognitive Theory (SCT).

Venkatesh et al. (2003) [27] have also tested the UTAUT resulting in a variance of 70%. Many scholars hence deem the UTAUT as the most comprehensive model in predicting employee behaviour in the e-government setting [12], [13], [29], [30], [38], [39]. Venkatesh et al. (2003) [27], Venkatesh et al. (2012) [28], and Taherdoost (2018) [48] report that, compared to other frameworks, the UTAUT boasts a better explanatory capability in forecasting the behavioural intention toward an IS. This study therefore employs the following constructs to complete the measurement:

### 1) Performance Expectancy (PEX)

The *Performance Expectancy* construct defines the degree of individual belief in which one feels that using the system will help improve his or her job performance [27], [45]. In general, Puspitarini & Ardhani (2022) [14] have established the evidence that PEX influence the usage of a technology. Further, in the e-government area, Yavwa and Twinomurinzi (2018) [33] failed to validate the connection between PEX and the usage behaviour. Yet, many scholars have successfully verified the correlation between PEX and usage behaviour [30]–[32], [34]–[36]. Mutaqin and Sutoyo (2020) [36] even prove that PEX stands as the most influential factor in the behavioural intention of e-government usage.

In the area of OGD-related study, Zainal et al. (2019) [39] propose an UTAUT-based model, which includes PEX, to measure the use of OGD in the academic setting. Further, Subedi et al. (2022) [26] certify the presence of PEX construct as the influential factor among OGD use by Nepali users. Hence, the dimension measured in this construct comprises *perceived usefulness, job-fit, and outcome expectations* [27], [28], [45] to gauge the acceptance of SIMDASI with the following hypothesis:

H8. PEX positively and significantly influences on SU to operate SIMDASI for OGD adoption.

### 2) Effort Expectancy (EEX)

The notion of the *Effort Expectancy* construct introduces a level of comfort related to using the accepted technology [27], [30], [45]. In the integrated model, Puspitarini & Ardhani (2022) [14] establish definitive proof that EEX affects the behavioural usage of the system. Albeit Taiwo et al. (2012) [34] dismissed the link between EEX and system usage, many

researchers propose evidence from their findings point to the connection between EEX and the usage of e-government services [30]–[33], [35], [36], [38].

In their research, Zainal et al. (2019) [39] expound on the correlation between EEX and system usage. In addition, Subedi et al. (2022) [26] depict the link between PEX and OGD use in Nepal. The authors thus propose this construct with the following measurement dimensions: *perceived ease of use, learning time, and complexity* [27], [28], [45]. For this dimension, the authors also formulate the hypothesis:

H9. EEX positively and significantly influences on SU to operate SIMDASI for OGD adoption.

### 3) Facilitating Conditions (FC)

The notion, known as the *Facilitating Conditions* construct, depicts how much one assumes the organisational and technological infrastructure support using the system. [27], [45]. Although Subedi et al. (2022) [26] ignore this construct, other scholars using the FC approach have presented the evidence. They claim that FC influences the behavioural use and the use of the system [30], [32]–[36], [39]. As for the measurement, this construct proposes the following criteria: *perceived behavioural control, facilitating materials, and compatibility*. Thus, the authors postulate the following hypothesis:

H10. FC positively and significantly influences on SU to operate SIMDASI for OGD adoption.

As for the *Social Influence* (SI), the authors believe that mandatory IS [8] compels employees to use the system without the influence of people around them. Al-Swidi & Faaeq (2019) [32] and Faroqi et al. (2020) [35] have proven that SI does not influence the usage behaviour in e-government services; the authors then decide to omit this construct from the model.

Finally, to unravel the benefits of SIMDASI in the context of OGD adoption. The authors formulate the following hypotheses:

H11. US positively and significantly influences on OI from SIMDASI as the OGD adoption tool.

H12. US positively and significantly influences on II from SIMDASI as the OGD adoption tool.

H13. SU positively and significantly influences on II from SIMDASI as the OGD adoption tool.

H14. SU positively and significantly influences on OI from SIMDASI as the OGD adoption tool.

Fig. 1 illustrates the measurement model in this study.

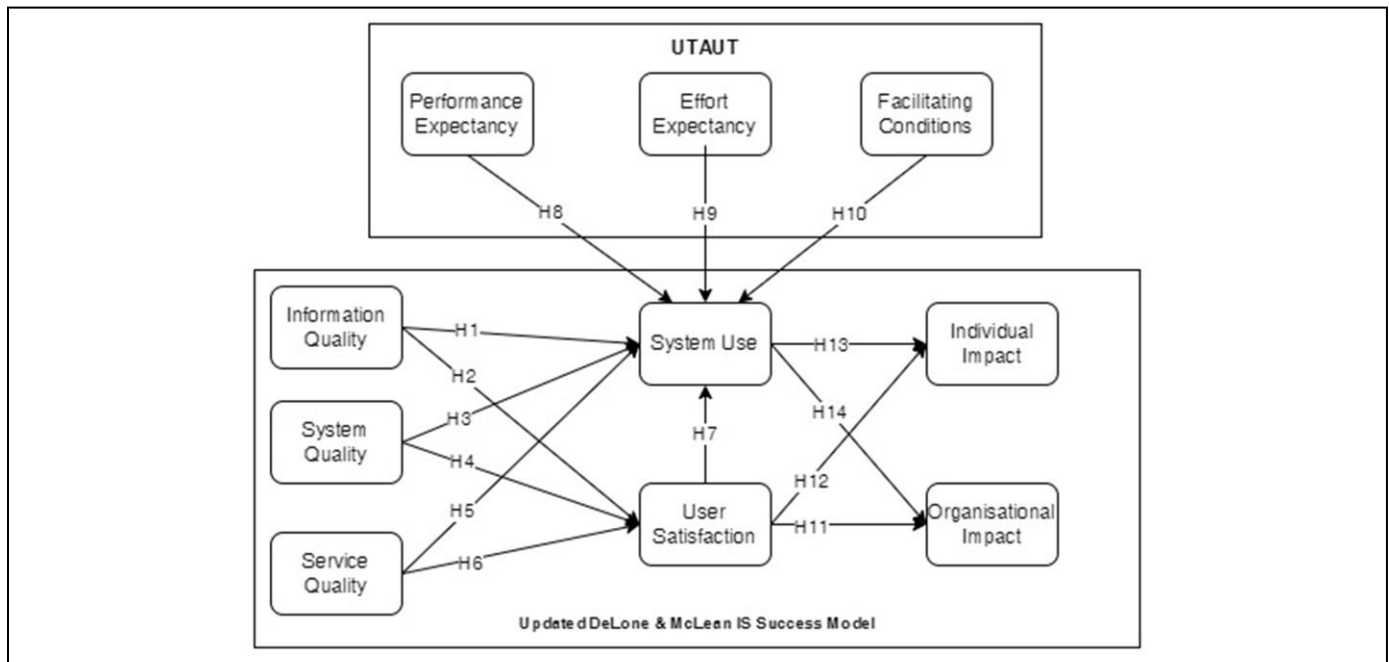


Fig. 1. Proposed measurement model

### III. RESEARCH METHODOLOGY

#### A. Design of the Study

This study outlines findings on the factors influencing OGD adoption through an integrated system based on employee views. The authors identify four endogenous constructs (system use, user satisfaction, organisational impacts, and individual impacts) and two mediating factors (user satisfaction and system use). Fig. 1 depicts the suggested model, which incorporates two models (ISSM and UTAUT), resulting in ten latent variables: *Information Quality (IQ)*, *Service Quality (SEQ)*, *System Quality (SQ)*, *User Satisfaction (US)*, *System Use (SU)*, *Performance Expectancy (PEX)*, *Effort Expectancy (EEX)*, *Facilitating Conditions (FC)*, *Individual Impacts (II)*, and *Organisational Impacts (OI)*.

The authors further establish fourteen different paths among the variables. In IS usage measurement, the authors construct seven hypotheses (H1–H7), while in uncovering employees' acceptance, the authors employ three hypotheses (H8–H10). Finally, the authors also hypothesise that SU and US are predictors of individual and organisational impact with four hypotheses for direct effects (H11–H14) [Fig. 1].

Furthermore, the authors employ a quantitative technique to obtain data, utilising questionnaires as the instrument. To pinpoint the correlation between one or more independent/dependent variables, the authors also use the structural equation model (SEM) methodology [49]. This study also focuses on BPS as the leading sector of OGD in Indonesia,

hence resulting in a *small sample size*. With addition of a *complex measurement model* in the study (many indicators and relationships), the authors hence select the partial least square (PLS) SEM as the modelling technique for the study [50]. Previous studies show that PLS SEM helped researchers in performing statistical analysis in a similar setting [13], [14], [16], [31], [32], [35], [43].

As for data analysis, the authors administer PLS software called SmartPLS 4 [51] to execute PLS-SEM [50], [52] and one-tailed bootstrapping tests [50], [53]. Finally, as advocated by Hair et al. (2017) [50], the authors utilise two-stage analysis: *measurement model assessment* and *structural model assessment*.

#### B. Instrumentation

The authors then collect the data using a cross-sectional questionnaire with close-ended questions. The questionnaire used in this study is designed in two parts: *respondents' information* and *respondents' view* regarding the proposed model. The first part of the questionnaire comprises eight questions that collect basic information about the respondents: *name, email, phone number, location (office), gender, age, education level, and working experience*. The second part contains 44 indicators to evaluate the proposed model. Each item is assessed on a five-point Likert scale: *1-strongly disagree; 2-disagree; 3-neutral; 4-agree; 5-strongly agree*. Appendix A exhibits the measurement items for this study and Fig. 2 depicts the recommended indicators for the model.

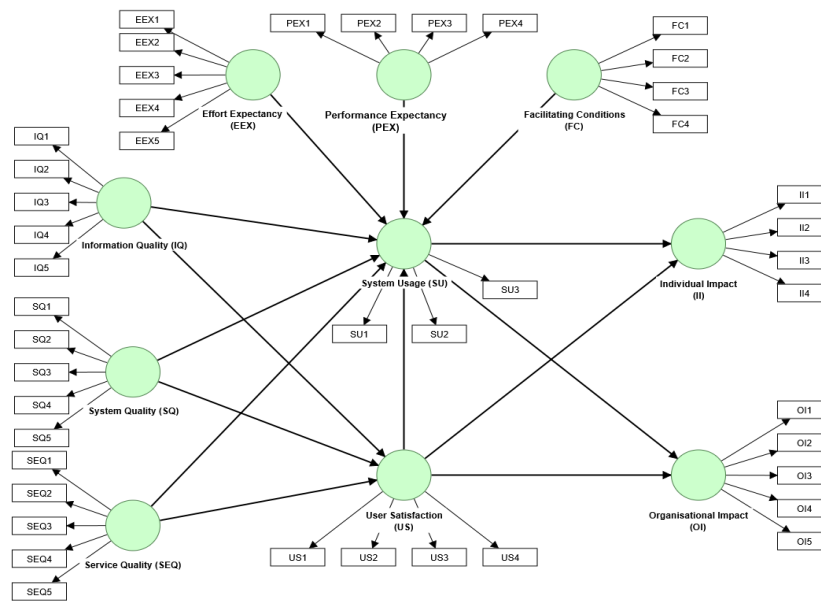


Fig. 2. Proposed measurement model with 44 indicators

C. Participants

The authors sent the questionnaire to approximately 613 employees who operate SIMDASI frequently. The authors then distribute through emails registered to SIMDASI from all BPS offices (headquarters, provincial offices, and regency/municipality offices), of which 253 responses were collected between 9 and 18 November 2022. In collecting the data, the authors inform and guarantee the security of the answer as their responses will remain secret and used only for academic purposes.

$$n = \frac{N}{1 + Ne^2} \quad (1)$$

To determine whether the collected responses have met the minimum sample requirement, the authors employ Slovin's formula to check the validity in equation (1), where *n* equals the required sample size, *N* means population size, and *e* is the margin of error. As Tejada et al. (2012) [54] recommended, the authors use a 95% confidence level or 5% of margin error, resulting in a 242 minimum sample size from a known 613 population. In a similar fashion, Kock & Hadaya (2018) [55] propose a sample size of 146 or 160 should researchers do not know the value of path coefficient with the minimum absolute magnitude. Thus, this study has met the required sample size.

Furthermore, as seen in Table I, 88,2% of the valid responses come from regional offices (province and regency/municipality offices), with most of the respondents being male (58,9 per cent). Then, nearly half of the respondents are 30–39 years of age (48,2 per cent). Then, nearly 99 per cent of respondents possessing higher education (diploma to doctorate) backgrounds and 39,5 per cent of the participants had worked for 11 to 20 years in BPS.

TABLE I. RESPONDENTS' PROFILE (N=253)

Item/Description	N	%
<b>Location (Office)</b>		
Headquarter	30	11.9
Provincial Office	66	26.1
Regency/Municipality Office	157	62.1
<b>Gender</b>		
Male	149	58.9
Female	104	41.4
<b>Age (years)</b>		
20–29	73	28.9
30–39	122	48.2
40–49	39	15.4
> 50	19	7.5
<b>Education Level</b>		
High School	1	0.4
Diploma	4	1.6
Bachelor's Degree	170	67.2
Master's Degree	77	30.4
Doctorate	1	0.4
<b>Working Experience (years)</b>		
< 5	51	20.2
5–10	68	26.9
11–20	100	39.5
> 20	34	13.4



#### IV. RESULT

##### A. Measurement Model Assessment

Hair et al. (2017) [50] and Hair et al. (2019) [52] recommend tests for examining the measurement model: *reflective indicator loadings, internal consistency, convergent validity, and discriminant validity.*

##### 1) Reflective indicator loadings (outer loadings)

Hair et al. (2017) [50] employ indicators of a reflective construct as a different technique for evaluating the same construct. The higher the outer loadings, the more common are the associated indicators on a construct. According to Hair et al. (2017) [50], an indicator's outer loading should be greater than 0.708 and, in most cases, 0.700 is adequate. Further, when an outer loading of an indicator is < 0.40, Hair et al. (2017) [50] recommend deleting the reflective indicator, while an outer loading between 0.40 and 0.70 should be analysed to check the increases of *average variance extracted (AVE)*.

From the calculation in Table II, two latent variables possess < 0.700 outer loading values, namely, IQ5 and SQ5. Accordingly, the authors prompted another reflective indicator loadings test by omitting **IQ5** and **SQ2** to examine changes in AVE. As identified in Table III, the AVE from IQ and SQ increase to 0.052 and 0.082, respectively. Thus, the authors agree to exclude IQ5 and SQ5 from the proposed model, resulting in above 0.700 outer loading values for all indicators shown in Table IV.

TABLE II. INITIAL REFLECTIVE OUTER LOADINGS

Indicator	Loading	Indicator	Loading
EEX1	0.826	OI5	0.879
EEX2	0.887	PEX1	0.852
EEX3	0.925	PEX2	0.915
EEX4	0.915	PEX3	0.935
EEX5	0.861	PEX4	0.903
FC1	0.840	SEQ1	0.917
FC2	0.792	SEQ2	0.924
FC3	0.843	SEQ3	0.941
FC4	0.789	SEQ4	0.888
II1	0.872	SEQ5	0.915
II2	0.909	SQ1	0.836
II3	0.931	SQ2	0.879
II4	0.907	SQ3	0.833
IQ1	0.748	SQ4	0.728
IQ2	0.800	<b>SQ5</b>	<b>0.608</b>
IQ3	0.848	SU1	0.830
IQ4	0.784	SU2	0.906
<b>IQ5</b>	<b>0.696</b>	SU3	0.870
OI1	0.865	US1	0.918

Indicator	Loading	Indicator	Loading
OI2	0.924	US2	0.882
OI3	0.879	US3	0.926
OI4	0.904	US4	0.914

TABLE III. CHANGES IN THE AVE BEFORE AND AFTER EXCLUSION

Latent Variable	AVE		Changes
	Before	After	
EEX	0.781	0.781	0.000
FC	0.666	0.666	0.000
II	0.819	0.819	0.000
<b>IQ</b>	<b>0.604</b>	<b>0.656</b>	<b>0.052</b>
OI	0.793	0.793	0.000
PEX	0.813	0.813	0.000
SEQ	0.842	0.842	0.000
<b>SQ</b>	<b>0.613</b>	<b>0.695</b>	<b>0.082</b>
SU	0.756	0.755	0.000
US	0.829	0.829	0.000

According to Table IV, Service Quality, SEQ3 obtains the most significant loading with 0.946, while the most negligible loading alluded to System Quality, SQ4 (0.718). After analysing the indicators, forty-two items were incorporated for the following assessment:

##### 2) Internal consistency reliability

The subsequent measurement to be examined is internal consistency reliability: *Cronbach's Alpha (Alpha)* and *Composite Reliability (CR)*. Alpha assumes that all indicators are equally reliable and are used to measure internal consistency reliability. Due to its limitations, Hair et al. (2017)[50] introduce an additional measure called CR, which accounts for different outer loadings of the indicators. Additionally, Hair et al. (2017) [50] and Hair et al. (2019) [52] emphasise that both CR and Alpha should be above 0.700.

Table VI shows that both Alpha and CR for all structures are acceptable, exceeding the appropriate level. For the acceptance factors (EEX, FC, PEX), the recorded Alphas are 0.929, 0.834, and 0.923, consecutively. As for the CR, these factors counted 0.947, 0.889, and 0.946, respectively. Regarding IS success, the measured factors (IQ, SQ, SEQ, US, SU, II, and OI) possess Alpha values in the following order: 0.825, 0.852, 0.953, 0.931, 0.840, 0.926, and 0.935. As for the CR values, the IS success factors recorded 0.884, 0.901, 0.964, 0.951, 0.902, 0.948, and 0.950.

##### 3) Convergent validity

Convergent validity constitutes a measure that correlates positively with alternative measures of the same dimension. Therefore, the indicators should converge, sharing a considerable proportion of variance. Hair et al. (2017) [50] state that the appropriate measure for convergent validity is AVE with a value above 0.500 or more, explaining 50% or

more of the variance. Table IV shows that all constructs received AVE values higher than the 0.500 variances.

4) *Discriminant validity*

Discriminant validity represents how a construct differs from others by empirical standards, implying the uniqueness among the variables. *The Fornell-Larcker criterion* and *heterotrait-monotrait ratio (HTMT)* are the best measurements to gauge the discriminant validity of the model [50], [52].

The Fornell-Larcker construct compares the square root of the AVE values with the latent variable correlations [50], [52]. The correlation between the two variables should be lower than the *square root of the AVE*. Appendix C exhibits this concept;

the square root of EEX AVE (0.781) is 0.884, and the correlation between FC and EEX is 0.750, which means that FC and EEX achieve discriminant validity.

Furthermore, HTMT translates as the ratio of the between-trait correlations of indicators among all variables. To obtain discriminant validity, the HTMT should be below 0.900. Appendix B depicts all HTMTs from the latent variables that are < 0.900 and vary from one another. Through the Fornell-Larcker criterion and HTMT examination, the authors hereby declare that the proposed model achieves discriminant validity.

TABLE IV. INDICATOR LOADINGS, INTERNAL CONSISTENCY RELIABILITY, AND CONVERGENT VALIDITY

Latent Variable	Item	Loading	Alpha	CR	AVE	Latent Variable	Item	Loading	Alpha	CR	AVE
Effort Expectancy (EEX)	EEX 1	0.826	0.929	0.947	0.781	Performance Expectancy (PEX)	PEX 1	0.852	0.923	0.946	0.813
	EEX 2	0.887					PEX 2	0.915			
	EEX 3	0.925					PEX 3	0.935			
	EEX 4	0.915					PEX 4	0.903			
	EEX 5	0.861									
Facilitating Conditions (FC)	FC1	0.840	0.834	0.889	0.666	Service Quality (SEQ)	SEQ 1	0.917	0.953	0.964	0.842
	FC2	0.792					SEQ 2	0.924			
	FC3	0.843					SEQ 3	0.941			
	FC4	0.789					SEQ 4	0.888			
							SEQ 5	0.915			
Individual Impact (II)	II1	0.872	0.926	0.948	0.819	System Quality (SQ)	SQ1	0.859	0.852	0.901	0.695
	II2	0.909					SQ2	0.909			
	II3	0.931					SQ3	0.838			
	II4	0.907					SQ4	0.718			
Information Quality (IQ)	IQ1	0.774	0.825	0.884	0.656	System Use (SU)	SU1	0.829	0.840	0.902	0.755
	IQ2	0.809					SU2	0.906			
	IQ3	0.869					SU3	0.871			
	IQ4	0.783									
Organisational Impact (OI)	OI1	0.865	0.935	0.950	0.793	User Satisfaction (US)	US1	0.918	0.931	0.951	0.829
	OI2	0.924					US2	0.882			
	OI3	0.879					US3	0.926			
	OI4	0.904					US4	0.914			
	OI5	0.879									

B. *Structural Model Assessment*

In assessing the structural model, Hair et al. (2017) [50] assert that PLS-SEM is beneficial in estimating the parameter to maximise the explained variance of the endogenous latent variable(s). The key criteria for evaluating the structural model are *collinearity*, *path coefficients*, and *coefficient of determination (R<sup>2</sup> values)* [50], [52].

1) *Collinearity*

Collinearity refers to the degree of correlation between two predictor constructs; thus, a high level of collinearity is crucial [50], [52]. To measure the collinearity, Hair et al. (2017) [50] recommend administering *variance inflation factor (VIF)*, and, in the context of PLS-SEM, the VIF value should be in a range between 0.20 and 5. Appendix D delineates the VIF statistics from the inner model and shows that all VIFs are within

recommended values. Therefore, collinearity is not a concern for the model used in this research.

2) Path coefficient ( $\beta$ )

Representing the hypothesised relationships among the constructs, the path coefficients ( $\beta$ ) possess standardised values between -1 and +1 [50], [52]—the closer  $\beta$  value is to 0, the weaker is the relationship. Then, to check whether the  $\beta$ s are significant, the authors attempt a one-tailed bootstrapping test procedure with 5,000 resamples, resulting in t-values and p-values. The authors compare the critical value and inspect the significance of every hypothesis.

Table VI informs the estimation results for the proposed hypotheses in this study. The path from user satisfaction to individual impact possesses the highest value ( $\beta=0.662$ ;  $t=10.508$ ), while the relationship from facilitating conditions to individual impact and organisational impact with 0.018 value for both paths. In addition, Table VI also confirms that eleven out of fourteen proposed hypotheses are supported.

3) Coefficient of determination ( $R^2$ )

$R^2$  value becomes the most applied measurement to assess the structural model by measuring the model's predictive capability [50], [52]. This approach allows researchers to examine the coefficient between exogenous latent variables' combined effects on the endogenous variable [50]. Table V explains that all endogenous model variables possess moderate effects [50], with individual impact receiving the highest  $R^2$ , meaning that system use and user satisfaction explain 71.1% of the variance. Moreover, information quality, system quality, service quality, and user satisfaction explain a 55.8% variance in system use, the lowest  $R^2$  value in the model.

TABLE V. COEFFICIENT OF DETERMINATION FROM THE MODEL

Endogenous Variable	$R^2$
Individual Impact	0.711
Organisational Impact	0.654
System Use	0.558
User Satisfaction	0.644

TABLE VI. ESTIMATION RESULTS FOR PROPOSED HYPOTHESES

Hypothesis	Relationship (from $\rightarrow$ to)	Path ( $\beta$ )	t-value	p-values	Result
H1	Information Quality $\rightarrow$ System Use	-0,003	0,033	0,487	Not Supported
H2	Information Quality $\rightarrow$ User Satisfaction	0,329	5,123	0,000	Supported*
H3	System Quality $\rightarrow$ System Use	0,163	2,013	0,022	Supported**
H4	System Quality $\rightarrow$ User Satisfaction	0,308	4,281	0,000	Supported*
H5	Service Quality $\rightarrow$ System Use	0,107	1,525	0,064	Supported***
H6	Service Quality $\rightarrow$ User Satisfaction	0,291	5,506	0,000	Supported*
H7	User Satisfaction $\rightarrow$ System Use	0,191	1,676	0,047	Supported**
H8	Performance Expectancy $\rightarrow$ System Use	0,242	2,953	0,002	Supported*
H9	Effort Expectancy $\rightarrow$ System Use	0,114	1,161	0,123	Not Supported
H10	Facilitating Conditions $\rightarrow$ System Use	0,059	0,659	0,255	Not Supported
H11	User Satisfaction $\rightarrow$ Organisational Impact	0,570	9,751	0,000	Supported*
H12	User Satisfaction $\rightarrow$ Individual Impact	0,604	9,367	0,000	Supported*
H13	System Use $\rightarrow$ Individual Impact	0,306	5,412	0,000	Supported*
H14	System Use $\rightarrow$ Organisational Impact	0,304	5,199	0,000	Supported*

Note: \* 1% significance level, \*\* 5% significance level, \*\*\* 10% significance level

V. DISCUSSION

The findings from this study provide insightful information into the factors influencing the success of IS as the mediator for OGD adoption. Through the integration of ISSM and UTAUT, the study successfully unravels the influencing factors in IS and acceptance manner from employees' perspective.

The proposed forty-four indicators establish ten latent variables developed. The authors exclude two indicators through outer loadings assessment, leading to valid constructs for further assessment. The forty-two indicators establish internal consistency reliability and validity through convergent and discriminant tests. In the structural model assessment, the

VIF values verify that no collinearity exists in the model. With the path coefficient test, the authors test the hypotheses at 1% and 5% significance levels. In addition, the authors also determine the  $R^2$  value to explain the model's predictive power. The authors then also investigate internal documents [8]–[10] for further analysis.

*RQ1: What are the IS factors that influence the use of SIMDASI?*

IQ provided by SIMDASI shows no sign of effects on SU from employees (H1) ( $\beta=-0.003$ ;  $t=0.033$ ); however, IQ possesses a significant correlation with US in operating SIMDASI (H2) ( $\beta=0.329$ ;  $t=5.123$ ). H2 resonates with the previous studies proving that IQ influences US [14], [16]. H1,

however, shows that IQ provided from SIMDASI has no influencing power in the use of SIMDASI. This finding differs from previous studies by Stefanovic et al. (2016) [15], Martono et al. (2020) [13], and Talukder et al. (2019) [43] that identified IQ as a salient construct affecting system use. The nature of employees working in data-driven departments [8]–[10] might cause the absence of a positive correlation between information quality and system use (H1), as employees are accustomed to working with all types of data and information quality.

Then, SQ exercised by SIMDASI resulted in a proven correlation to SU (H3) ( $\beta = 0.163$ ;  $t = 2.013$ ) and a causal effect to US (H4) ( $\beta = 0.308$ ;  $t = 4.281$ ). H3 and H4 further prove the same results from Puspitarini & Ardhani (2022) [14], Talukder et al. (2019) [43], Stefanovic et al. (2016) [15], Martono et al. (2020) [13] and Gangga Dewi & Fajar (2021) [16]. This research shows that the system's quality is a driving factor leading to more usage and higher satisfaction among employees in BPS. To put differently, employees are keen to use the system when it offers adequate usability and performance.

SEQ provided by the department in charge of SIMDASI has significant connections for both SU (H5) and US (H6) with  $\beta$  values of 0.107 and 0.291 with t-value of 1.525 and 5.506, consecutively. This outcome further proves the notion from Puspitarini & Ardhani (2022) [14] and Gangga Dewi & Fajar (2021) [16]. The presence of service personnel providing a hotline and helpdesk [8] becomes the main reason in this finding. The findings prove that aid from service personnel becomes paramount in the use of the system and the employees' happiness, corresponding to the original theory from ISSM [44]–[46].

Table VI also confirms the existence of the US-SU correlation (H7) with  $\beta$  value of 0.191 and a t-value of 1.676. This finding reaffirms the evidence from Gangga Dewi & Fajar (2021) [16] and Stefanovic et al. (2016) [15]. The results above prove that only IQ, SQ, and SEQ influence the employees' happiness in operating SIMDASI. Then, three IS-based constructs remain influential regarding using SIMDASI: SQ, SEQ, and US.

In view of the above, the authors find that the quality of system and services, along with satisfaction among employees, become the influencing factor from SIMDASI, which affects its usage. In addition, IQ, SQ, and SEQ are the predictor of US, which subsequently affects SU.

*RQ2: What acceptance factors from employees affect the use of SIMDASI?*

In the acceptance factors, Table VI validates that PEX plays a significant role in determining system use (H8) with  $\beta$  value of 0.242 and a t-value of 2.953. This finding matches the same result from previous studies [30]–[32], [34]–[36]. This finding depicts that when employees perceive SIMDASI as valuable and influential to their job performance, their intention to use it increases, corresponding to the original theoretical foundation of UTAUT [27], [28].

On the other hand, EEX (H9) and FC (H10), surprisingly, are proven to be noninfluential for the use of SIMDASI. H9 ratifies the same result from Taiwo et al. (2012) [34], of which

they find no correlation between EEX and system usage. The findings elaborate that employees in BPS presume that the usage of SIMDASI requires effort (H9) and lacks support by an organisation and technological infrastructure (H10), reflected in the internal evaluation documents (such as lack of knowledge transfer, incomprehensive training, and medium to low participation from regional areas) [9], [10]. This study proves that the increase of effort expectancies or facilitating conditions plays no part in the increase of system use. The mandatory of SIMDASI could also lead to this finding, as the employees have no choice but to use it to disseminate OGD in BPS.

Hence, this study proves that performance expectancies remain the only independent variable from the acceptance factors that influence the use of SIMDASI.

*RQ3: What direct factors determine the success of SIMDASI as an OGD adoption tool?*

For the direct effect, US and SU are proven to be strong predictors for II and OI at a 1% significance level. The US in this model possesses the two highest path coefficient values in determining the net benefits: 0.604 (H12) of  $\beta$  to II and 0.570 of  $\beta$  to OI (H11)—denoting the findings from Stefanovic et al. (2016) [15] and Gangga Dewi & Fajar (2021) [16]. The use of the system and employees' happiness impact OGD adoption in individual and organisational contexts and empower employees and organisations via the adoption of OGD. This finding also corresponds to the updated theoretical foundation of ISSM [17], [46] that explains the predictors from II and OI are SU and US.

This study therefore concludes that system use and employees' satisfaction become the predictor variables in determining the success of SIMDASI toward OGD initiatives in BPS. Those two variables account for 71.11% of the variation in II and 65.4% of the variance in OI.

This research contributes to the literature by proposing an integrated model while also demonstrating a valuable paradigm for understanding the following:

- Information Quality, Service Quality, and System Quality strongly influence employees' satisfaction in operating SIMDASI.
- Service quality, system quality, and user satisfaction become influential in escalating the use of SIMDASI among employees.
- Performance expectancies may also improve employees' usage of SIMDASI.
- The rise in usage of SIMDASI and employees' satisfaction leads to positive development for individuals and organisations alike.

#### A. Theoretical Implications

The proposed model comprises seven constructs, namely, IQ, SQ, SEQ, US, PEX, EEX, and FC are used as potential dimensions that may influence the use of SIMDASI. In comparison, three constructs are established to explain user satisfaction: IQ, SQ, and SEQ. US and SU constructs are then

evaluated to examine the influential power for personal and organisational impact from SIMDASI to OGD initiative.

This research employs two frameworks (ISSM and UTAUT) to assess the findings. The significant theoretical impacts and implications of this research are elaborated below:

- The ISSM model has been proven to be an ideal framework for understanding Information Quality, Service Quality, and System Quality in enhancing employee satisfaction (H2, H4, and H6), which subsequently could improve the use of the system (H7). This model also helps explain the predictor variables from individual and organisational impacts: user satisfaction and system use (H11–H14).
- The UTAUT model has offered evidence that performance expectancies (H8) play an increasingly pivotal role in using SIMDASI as the medium of individual and organisational benefits. However, in a mandatory setting, effort expectancy and facilitating conditions become less influential for system use (H9 and H10).
- The integrated model covers the weaknesses of each model, resulting in a better predictive power to uncover the benefits of IS for an organisation from the employees' perspective.

#### B. Practical Implications

This study possesses a wide array of practical implications for decision-makers in the government. This study also promotes a variety of priorities to comprehend employees' acceptance of using IS and strengthen employees' happiness with operating the IS by combining two ideas of IS and acceptance: ISSM and UTAUT. Although some hypotheses regarding acceptance constructs are not supported, one of the acceptance factors, such as PEX, holds predictive power in explaining the increase in system use. PEX subsequently influences the individual and organisational benefits through the medium of system use.

The following are the key findings of the study in a practical manner:

- In increasing satisfaction among employees, decision-makers should be more mindful of the quality of the information provided by the system (usefulness, understandable, engaging, and reliable), the availability of services to aid the employees (willingness, personal attention, punctuality, resourceful, and completeness), and the performance of the proposed system (easy to use and navigate, interactive, and accessible).
- With the increase in employee satisfaction, service quality, and system quality, the decision-makers have the upper hand in enhancing the use of the system; thus, all involved parties must attend to this matter.
- Decision-makers should also establish employees' beliefs regarding how the IS may improve their job performance. This perception may alter the use of the proposed system in the process. Decision-makers should attend to the resources and knowledge available

for the employees, the compatibility with other technologies, and the presence of a specific person for assistance.

- Government ministries/agencies could benefit from an integrated IS should they adopt OGD. Consequently, the adopters must pay close attention to the satisfaction of employees as they operate the system on a frequent basis.

#### VI. CONCLUSION

This study aims to pinpoint factors contributing to the success of an integrated IS to OGD initiative from the employees' perspective. The authors declare the objectives for this research: (1) to investigate the IS factors affecting the use of SIMDASI in OGD initiative; (2) to examine acceptance factors in the usage of SIMDASI; (3) to uncover influencing factors from the success of SIMDASI employees' point of view in adopting OGD. Thus, with the integration of ISSM and UTAUT, this paper covers the IS success factors and user acceptance by measuring ten different latent variables: IQ, SQ, SEQ, US, PEX, EEX, FC, SU, II, and OI. The authors then postulate twenty-six hypotheses to answer the objectives.

The authors then administer a quantitative method using a cross-sectional questionnaire with close-ended questions. Between 9 and 18 November 2022, the researchers collected 253 responses from the employees in BPS who use SIMDASI. The authors also decide to employ SEM in determining the correlations among the latent variables with the help of SmartPLS 4.

Moreover, this paper examines the validity and reliability of the proposed measurement model of ISSM-UTAUT through measurement model assessment. The test resulted in the exclusion of two indicators among 44 initial indicators. Reflective indicator loadings, internal consistency reliability, convergent validity, and discriminant validity—these are the tests conducted to confirm the model's validity. In addition, this paper also conducts a test on the structural model to verify the proposed hypotheses using the path coefficient. Eleven out of fourteen hypotheses are supported and the additional test of  $R^2$  values presents comprehensive information about the variance caused by exogenous factors.

The accepted hypotheses give insight into the research objectives. Although information quality influences employees' satisfaction in operating the system, information quality remains noninfluential for the usage of the system. However, the quality of the system and the services (provided by a specific department) are proven to affect system use and employees' happiness.

Regarding the acceptance factors, employees' belief in using the system to improve their job performance (PEX) has become the driving factor leading to the acceptance of SIMDASI. Nevertheless, this paper disproves the connection between employees' degree of ease (EEX) associated with SIMDASI and the use of SIMDASI. The same goes for the organisational and technical infrastructure support (FC) plays no part in defining the use of SIMDASI.

In general, this paper uncovers the influencing factors from the benefits generated by SIMDASI to OGD initiatives: individual and organisational benefits. The outcome of the assessment shows that system use and user satisfaction are predictors of the benefits from the view of IS. The result also clarifies that only performance expectancy has a part in defining the benefits of SIMDASI to OGD initiatives. Therefore, the use of system and employees' happiness stand as the explanatory variables for organisational and individual impact in OGD initiatives.

#### A. Limitations of Study

Albeit this study presents exciting findings, it has certain limitations: *participants and the measurement model*.

First, is the limitation of sample size and the focus on a single type of organisation. As the Central Data Supervisory Institution in Statistics, BPS owns a better understanding of disseminating statistical data; hence the OGD adoption might be acceptable. Additionally, with relatively supporting policies available, adopting new technologies becomes more straightforward. Thus, the findings of this study may reflect different factors in other government ministries/agencies.

Second, is the measurement model. The measurement model lacks specificity for the observed area by only integrating two predefined models. Other factors that might contribute to the success of IS in adopting OGD such as experience/habit, characteristics of technology and task, or employees' training prior to the use.

Finally, despite the limitations, this study successfully pinpoint the influencing factors using IS to OGD adoption.

#### B. Future Work

The authors suggest expanding the sample size and more coverage from government organisations such as ministries/agencies from central to local. Moreover, employing a longitudinal approach may result in more accurate findings to explain the behaviour of the employees. The data collection method could be improved by administering an assisted interview method to help respondents understand the questions better.

Future research could also exercise the same measurement model in different backgrounds: different countries or types of organisations, thus, validating the extent of the model used in this study. Additionally, since this study employs ISSM and UTAUT constructs, future research could extend this model by updating the model with the current version, such as UTAUT2. Next, introducing novel factors outside ISSM and UTAUT construct could provide a broader understanding of how IS influences the success of OGD adoption from employees' perspective. Task-Technology Fit (TTF) or Fit-Viability theory could help explain the impact of an integrated IS for government ministers/agencies in OGD initiatives.

Therefore, the authors believe that many possibilities are available in the future regarding this research—A better way to understand how an information system affects the adoption of open government data.

#### AUTHORS' CONTRIBUTIONS

WSW collects the data, performs data curation and statistical analysis, validates the research methodology, interprets the result, and writes the original manuscript. AF conducts research conceptualisation, interprets the data, and writes the original draft. DIA, SL, PAWP, and AY supervise the study, validate the results from WSW and AF, revise the mistakes, and review the original manuscript. All writers have reviewed and approved the final manuscript.

#### CONFLICT-OF-INTEREST STATEMENT

The authors with this declare that there is no conflict of interest in this study. All co-authors have reviewed and authorised the manuscript's contents; hence, the authors certify that the submission is authentic and is not currently under consideration by another publisher.

#### ACKNOWLEDGMENT

The authors would like to express their deepest gratitude to the Ministry of Communication and Information Technology (KOMINFO) for financial aid in carrying out this research. WSW and AF would also offer sincere appreciation to KOMINFO for supporting them during their study at Universitas Indonesia

#### REFERENCES

- [1] J. Tauberer, *Open Government Data: Second Edition*. 2014. [Online]. Available: <https://www.amazon.com/Open-Government-Data-Joshua-Tauberer-ebook/dp/B007U69DKU>
- [2] K. Okamoto, "Introducing Open Government Data," *Ref. Libr.*, vol. 58, no. 2, pp. 111–123, 2017, doi: 10.1080/02763877.2016.1199005.
- [3] H. Huang, C. Z. P. Liao, H. C. Liao, and D. Y. Chen, "Resisting by workarounds: Unraveling the barriers of implementing open government data policy," *Gov. Inf. Q.*, vol. 37, no. 4, p. 101495, 2020, doi: 10.1016/j.giq.2020.101495.
- [4] Open Knowledge Foundation, "Why Open Data?," 2015. <https://opendatahandbook.org/guide/en/why-open-data/> (accessed Sep. 11, 2022).
- [5] E. F. Ramos, "OPEN DATA DEVELOPMENT OF COUNTRIES: GLOBAL STATUS AND TRENDS," *ITU Kaleidosc. Acad. Conf.*, pp. 21–28, 2017.
- [6] United Kingdom's Government, "Open Standards principles," 2018. <https://www.gov.uk/government/publications/open-standards-principles/open-standards-principles> (accessed Sep. 11, 2022).
- [7] President of Indonesia, *Peraturan Presiden Republik Indonesia Nomor 39 Tahun 2019 Tentang Satu Data/ the Republic of Indonesia Presidential Decree Number 39 of 2019 about One Data*. 2019.
- [8] Badan Pusat Statistik, "SIMDASI Manual: Integrated Statistical Data Management Information System/Manual SIMDASI: Sistem Informasi Manajemen Data Statistik Terintegrasi," 2019.
- [9] Badan Pusat Statistik, "Evaluation Report and Achievements from the Implementation of SIMDASI/ Laporan Evaluasi dan Capaian Implementasi SIMDASI," 2020.
- [10] Badan Pusat Statistik, "Evaluation of SIMDASI/ Evaluasi SIMDASI," 2021.
- [11] R. Gholami, N. Singh, P. Agrawal, K. Espinosa, and D. Bamufleh, "Information Technology/Systems Adoption in the Public Sector," *J. Glob. Inf. Manag.*, vol. 29, no. 4, pp. 172–194, Jul. 2021, doi: 10.4018/JGIM.20210701.oa8.
- [12] A. W. Nur Syuhaini and B. László, *A Proposed Model for Assessing E-Government Adoption Among Civil Servants*, vol. 1, no. 1. Association for Computing Machinery, 2022. doi: 10.1145/3551504.3551545.

- [13] S. MARTONO, A. NURKHIN, H. MUKHIBAD, I. ANISYKURLILLAH, and C. W. WOLOR, "Understanding the Employee's Intention to Use Information System: Technology Acceptance Model and Information System Success Model Approach," *J. Asian Financ. Econ. Bus.*, vol. 7, no. 10, pp. 1007–1013, 2020, doi: 10.13106/jafeb.2020.vol7.no10.1007.
- [14] A. Puspitarini and A. R. Ardhani, "EXTENDED DELONE & MCLEAN ISS MODEL TO EVALUATE IT ASSISTANCE APPLICATION USAGE LEVEL," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 19, 2022.
- [15] D. Stefanovic, U. Marjanovic, M. Delić, D. Culibrk, and B. Lalic, "Assessing the success of e-government systems: An employee perspective," *Inf. Manag.*, vol. 53, no. 6, pp. 717–726, Sep. 2016, doi: 10.1016/j.im.2016.02.007.
- [16] A. A. I. K. Gangga Dewi and A. N. Fajar, "ASSESSING THE SUCCESS OF KOMINFO MAIL HANDLING SYSTEM BASED ON EMPLOYEE PERSPECTIVE," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 11, 2021.
- [17] W. H. DeLone and E. R. Mclean, "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update," *J. Manag. Inf. Syst.*, vol. 19, 2003, doi: <https://doi.org/10.1080/07421222.2003.11045748>.
- [18] D. Stefanovic, A. Milicevic, S. Havzi, T. Lolic, and A. Ivic, "Information Systems Success Models in the E-Government: Context: A Systematic Literature Review," in 2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH), Mar. 2021, no. March, pp. 1–6. doi: 10.1109/INFOTEH51037.2021.9400653.
- [19] H. N. Roa, E. Loza-Aguirre, and P. Flores, "A Survey on the Problems Affecting the Development of Open Government Data Initiatives," in 2019 Sixth International Conference on eDemocracy & eGovernment (ICEDEG), 2019, pp. 157–163. doi: 10.1109/ICEDEG.2019.8734452.
- [20] D. Donald Shao and S. Saxena, "Barriers to Open Government Data (OGD) initiative in Tanzania: Stakeholders' perspectives," *Growth Change*, vol. 50, no. 1, pp. 470–485, 2019, doi: 10.1111/grow.12282.
- [21] J. Wiczorkowski, "Barriers to using open government data," *ACM Int. Conf. Proceeding Ser.*, pp. 15–20, 2019, doi: 10.1145/3340017.3340022.
- [22] I. Safarov, "Institutional Dimensions of Open Government Data Implementation: Evidence from the Netherlands, Sweden, and the UK," *Public Perform. Manag. Rev.*, vol. 42, no. 2, pp. 305–328, 2019, doi: 10.1080/15309576.2018.1438296.
- [23] E. Shepherd et al., "Open government data: critical information management perspectives," *Rec. Manag. J.*, vol. 29, no. 1–2, pp. 152–167, 2019, doi: 10.1108/RMJ-08-2018-0023.
- [24] V. Wang, D. Shepherd, and M. Button, "The barriers to the opening of government data in the UK: A view from the bottom," *Inf. Polity*, vol. 24, no. 1, pp. 59–74, Mar. 2019, doi: 10.3233/IP-180107.
- [25] D. W. Jacob, M. F. M. Fudzee, M. A. Salamat, and N. H. A. Rahman, "Analyzing the barrier to open government data (OGD) in Indonesia," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1.3S1, pp. 136–139, 2019, doi: 10.30534/ijatcse/2019/2681.32019.
- [26] R. Subedi, T. E. Nyamasvisva, and M. Pokharel, "An Integrated-Based Framework For Open Government Data Adoption In Kathmandu," vol. 19, no. 2, pp. 7936–7961, 2022.
- [27] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User Acceptance of Information Technology: Toward a Unified View," *MIS Q. Manag. Inf. Syst.*, vol. 27, no. 3, pp. 425–478, 2003.
- [28] V. Venkatesh, J. Y. L. Thong, and X. Xu, "Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology," *MIS Q. Manag. Inf. Syst.*, vol. 36, no. 1, pp. 157–178, 2012.
- [29] T. Lolic, V. Ilic, S. Ristic, D. C. Lalic, and S. Havzi, "Exploring factors influencing the e-learning system success: a student perspective," no. March, pp. 16–18, 2022.
- [30] A. Sivaji, S. H. Rasidi, S. H. Hashim, A. D. N. Kuppasamy, and F. Z. Abidin, "Unified Theory of Acceptance and Use of Technology of E-Government Services in Malaysia: Validation of Survey Instrument," pp. 6–11, 2019.
- [31] M. M. Alzubi, B. Ismaeel, and A. H. Ateik, "The moderating effect of compatibility factor in the usage of E-government services among Malaysian citizens," 2021 2nd Int. Conf. Smart Comput. Electron. Enterp. Ubiquitous, Adapt. Sustain. Comput. Solut. New Norm. ICSCCE 2021, no. 2005, pp. 224–232, 2021, doi: 10.1109/ICSCCE50312.2021.9498076.
- [32] A. K. Al-Swidi and M. K. Faaeq, "How robust is the UTAUT theory in explaining the usage intention of e-government services in an unstable security context?: A study in Iraq," *Electron. Gov.*, vol. 15, no. 1, pp. 37–66, 2019, doi: 10.1504/EG.2019.096580.
- [33] Y. Yavva and H. Twinomurizi, "Impact of Culture on E-Government Adoption Using UTAUT: A Case of Zambia," 2018 5th Int. Conf. eDemocracy eGovernment, ICEDEG 2018, pp. 356–360, 2018, doi: 10.1109/ICEDEG.2018.8372350.
- [34] A. A. Taiwo, A. K. Mahmood, and A. G. Downe, "User acceptance of eGovernment: Integrating risk and trust dimensions with UTAUT model," 2012 Int. Conf. Comput. Inf. Sci. ICCIS 2012 - A Conf. World Eng. Sci. Technol. Congr. ESTCON 2012 - Conf. Proc., vol. 1, pp. 109–113, 2012, doi: 10.1109/ICCISci.2012.6297222.
- [35] A. Faraq, T. L. M. Suryanto, and E. M. Safitri, "The Determinant of E-Government Services Adoption among Citizen in Indonesia," *Proceeding - 6th Inf. Technol. Int. Semin. ITIS 2020*, pp. 130–134, 2020, doi: 10.1109/ITIS50118.2020.9321068.
- [36] K. A. Mutaqin and E. Sutoyo, "Analysis of Citizens Acceptance for e-Government Services in Bandung, Indonesia: The Use of the Unified Theory of Acceptance and Use of Technology (UTAUT) Model," vol. 1, no. 1, pp. 19–25, 2020, doi: 10.25008/bcsee.v1i1.3.
- [37] N. Z. Zainal, H. Hussin, and M. N. Mior Nazri, "Acceptance, Quality and Trust Factors - Conceptual Model for Open Government Data Potential Use," *Int. J. Perceptive Cogn. Comput.*, vol. 5, no. 2, pp. 12–18, 2019, doi: 10.31436/ijpcc.v5i2.120.
- [38] B. Soledad Fabito, R. L. Rodriguez, A. O. Trillanes, J. I. G. Lira, D. Z. Estocada, and P. M. Q. Sta Ana, "Investigating the Factors influencing the Use of a Learning Management System (LMS): An Extended Information System Success Model (ISSM)," in 2020 The 4th International Conference on E-Society, E-Education and E-Technology, Aug. 2020, pp. 42–46. doi: 10.1145/3421682.3421687.
- [39] N. Z. Zainal, H. Hussin, N. H. A. Rahim, M. N. M. Nazri, and M. A. Suhaimi, "Open Government Data Use by Malaysian Researchers. Some empirical evidence," in 2019 6th International Conference on Research and Innovation in Information Systems (ICRIIS), Dec. 2019, pp. 1–6. doi: 10.1109/ICRIIS48246.2019.9073640.
- [40] N. Z. Zainal, H. Hussin, and M. N. Mior Nazri, "A Trust-Based Conceptual Framework on Open Government Data Potential Use," *Int. Conf. Inf. Commun. Technol. Muslim World*, pp. 156–161, 2018, doi: 10.1109/ICT4M.2018.00037.
- [41] W. R. Fitriani, A. N. Hidayanto, and P. I. Sandhyaduhita, "Determinants of Continuance Intention to Use Open Data Website: An Insight from Indonesia," vol. 11, no. 2, pp. 96–120, doi: 10.17705/1pais.11205.
- [42] D. Krismawati and A. N. Hidayanto, "The User Engagement of Open Data Portal," 2021 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSIS 2021, 2021, doi: 10.1109/ICACSIS53237.2021.9631357.
- [43] M. S. Talukder, L. Shen, M. F. Hossain Talukder, and Y. Bao, "Determinants of user acceptance and use of open government data (OGD): An empirical investigation in Bangladesh," *Technol. Soc.*, vol. 56, pp. 147–156, 2019, doi: <https://doi.org/10.1016/j.techsoc.2018.09.013>.
- [44] W. H. DeLone and E. R. McLean, "Information Systems Success: The Quest for the Dependent Variable," *Inf. Syst. Res.*, vol. 3, no. 1, pp. 60–95, Mar. 1992, doi: 10.1287/isre.3.1.60.
- [45] Y. K. Dwivedi, M. R. Wade, and S. L. Schneberger, Eds., *Information Systems Theory*, vol. 29. New York, NY: Springer New York, 2012. doi: 10.1007/978-1-4419-9707-4.
- [46] W. H. DeLone and E. R. Mclean, "Information Systems Success Measurement," vol. 2, no. 1, pp. 1–116, 2016, doi: 10.1561/2900000005.
- [47] N. Urbach and B. Müller, "The Updated DeLone and McLean Model of Information Systems Success," 2012, pp. 1–18. doi: 10.1007/978-1-4419-6108-2\_1.
- [48] H. Taherdoost, "A review of technology acceptance and adoption models and theories," *Procedia Manuf.*, vol. 22, pp. 960–967, 2018, doi: 10.1016/j.promfg.2018.03.137.

- [49] J. B. Ullman and P. M. Bentler, "Structural Equation Modeling," in Handbook of Psychology, Second Edition, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2012. doi: 10.1002/9781118133880.hop202023.
- [50] J. F. Hair, G. T. Hult, C. Ringle, and M. Sarstedt, A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). 2017.
- [51] C. M. Ringle, S. Wende, and J.-M. Becker, "SmartPLS 4. Boenningstedt: SmartPLS.," 2022. <https://www.smartpls.com/> (accessed Nov. 11, 2022).
- [52] J. F. Hair, J. J. Risher, M. Sarstedt, and C. M. Ringle, "When to use and how to report the results of PLS-SEM," Eur. Bus. Rev., vol. 31, no. 1, pp. 2–24, Jan. 2019, doi: 10.1108/EBR-11-2018-0203.
- [53] N. Kock, "One-tailed or two-tailed P values in PLS-SEM?," Int. J. e-Collaboration, vol. 11, no. December, pp. 1–7, 2014.
- [54] J. J. Tejada, J. Raymond, and B. Punzalan, "On the Misuse of Slovin's Formula," Philipp. Stat., vol. 61, no. 1, p. 8, 2012.
- [55] N. Kock and P. Hadaya, "Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods," pp. 1–37, 2018.
- [56] M. A. Almaiah et al., "Factors influencing the adoption of internet banking: An integration of ISSM and UTAUT with price value and perceived risk," Front. Psychol., vol. 13, 2022, doi: 10.3389/fpsyg.2022.919198.

APPENDIX A: MEASUREMENT ITEMS

Dimension	Code	Items	Reference
<b>Information Quality (IQ)</b>	<b>IQ1</b>	The information provided by SIMDASI is useful.	
	<b>IQ2</b>	The information provided by SIMDASI is easy to understand	
	<b>IQ3</b>	The information provided by SIMDASI is interesting	[15], [17], [44], [46], [47], [56]
	<b>IQ4</b>	The information provided by SIMDASI is reliable.	
	<b>IQ5</b>	The information provided by SIMDASI is up to date	
<b>System Quality (SQ)</b>	<b>SQ1</b>	SIMDASI is easy to use	
	<b>SQ2</b>	SIMDASI is easy to navigate	
	<b>SQ3</b>	SIMDASI provides interactive features	[15], [17], [44], [46], [47]
	<b>SQ4</b>	SIMDASI is accessible	
	<b>SQ5</b>	SIMDASI provides integration with other systems	
<b>Service Quality (SEQ)</b>	<b>SEQ1</b>	The responsible service personnel are always highly willing to help whenever I need support with SIMDASI.	
	<b>SEQ2</b>	The responsible service personnel provide personal attention when I experience problems with SIMDASI.	[17], [44], [46], [47], [56]
	<b>SEQ3</b>	The responsible service personnel provide services related to SIMDASI at the promised time.	
	<b>SEQ4</b>	The responsible service personnel have sufficient knowledge to answer my questions regarding SIMDASI.	
	<b>SEQ5</b>	The SIMDASI overall service quality from assigned personnel is complete	
<b>System Use (SU)</b>	<b>SU1</b>	I spend 3 to 4 days weekly on SIMDASI to complete my task	
	<b>SU2</b>	I try new features and functions in SIMDASI for specific tasks to make me more efficient than others	[17], [44], [46], [47]
	<b>SU3</b>	I feel at ease when using SIMDASI	
<b>Facilitating Condition (FC)</b>	<b>FC1</b>	I have the resources necessary to use SIMDASI.	
	<b>FC2</b>	I have the knowledge necessary to use SIMDASI.	[27], [28], [45], [56]
	<b>FC3</b>	SIMDASI is compatible with other technologies I use.	
	<b>FC4</b>	A specific person is available for assistance with SIMDASI difficulties.	
<b>Performance Expectancy (PEX)</b>	<b>PEX1</b>	SIMDASI would be useful for me to complete my tasks.	
	<b>PEX2</b>	SIMDASI would allow me to complete my tasks more quickly.	[27], [28], [45], [56]
	<b>PEX3</b>	Using SIMDASI would increase my productivity levels.	
	<b>PEX4</b>	Using SIMDASI would improve my performance.	
<b>Effort Expectancy (EEX)</b>	<b>EEX1</b>	It would be easy for me to become skilful at using SIMDASI.	
	<b>EEX2</b>	Learning how to use SIMDASI is easy for me.	
	<b>EEX3</b>	My interaction with SIMDASI would be clear and understandable.	[27], [28], [45], [56]
	<b>EEX4</b>	I would find it easy to get SIMDASI to do what I want it to do.	
	<b>EEX5</b>	Overall, I believe that it is easy to use SIMDASI to support my tasks.	
<b>User Satisfaction (US)</b>	<b>US1</b>	I was very content with SIMDASI.	[15], [17],



	<b>US2</b>	I was very satisfied with the information in SIMDASI.	[44], [46], [47]
	<b>US3</b>	I was satisfied with the efficiency of SIMDASI.	
	<b>US4</b>	Overall, I felt delighted with SIMDASI.	
<b>Individual Impact (II)</b>	<b>II1</b>	I have learnt much through the presence of SIMDASI.	
	<b>II2</b>	SIMDASI enhances my awareness and recall of job-related information.	[15], [17], [44], [46], [47]
	<b>II3</b>	SIMDASI enhances my effectiveness in the job.	
	<b>II4</b>	SIMDASI increases my productivity.	
<b>Organisational Impact (OI)</b>	<b>OI1</b>	SIMDASI has resulted in cost reductions (e.g., administration expenses or data collection activities).	
	<b>OI2</b>	SIMDASI has resulted in overall productivity improvement in BPS.	
	<b>OI3</b>	SIMDASI has resulted in improved outcomes or outputs (e.g., data quality).	[17], [44], [46], [47]
	<b>OI4</b>	SIMDASI has resulted in improved business processes.	
	<b>OI5</b>	SIMDASI has resulted in better positioning for Open Data in BPS.	

APPENDIX B: DISCRIMINANT VALIDITY ANALYSIS MATRIX (HETERO TRAIT-MONOTRAIT RATIO)

	<b>EEX</b>	<b>FC</b>	<b>II</b>	<b>IQ</b>	<b>OI</b>	<b>PEX</b>	<b>SEQ</b>	<b>SQ</b>	<b>SU</b>	<b>US</b>
<b>EEX</b>										
<b>FC</b>	0.849									
<b>II</b>	0.801	0.760								
<b>IQ</b>	0.789	0.751	0.664							
<b>OI</b>	0.713	0.726	0.880	0.690						
<b>PEX</b>	0.773	0.746	0.816	0.723	0.772					
<b>SEQ</b>	0.656	0.639	0.666	0.622	0.608	0.606				
<b>SQ</b>	0.742	0.698	0.685	0.801	0.668	0.598	0.690			
<b>SU</b>	0.722	0.685	0.800	0.661	0.768	0.727	0.636	0.693		
<b>US</b>	0.819	0.758	0.874	0.796	0.833	0.799	0.707	0.798	0.757	

APPENDIX C: DISCRIMINANT VALIDITY ANALYSIS MATRIX (FORNER-LARCKER CRITERION)

	<b>EEX</b>	<b>FC</b>	<b>II</b>	<b>IQ</b>	<b>OI</b>	<b>PEX</b>	<b>SEQ</b>	<b>SQ</b>	<b>SU</b>	<b>US</b>
<b>EEX</b>	<b>0.884</b>									
<b>FC</b>	0.750	<b>0.816</b>								
<b>II</b>	0.748	0.676	<b>0.905</b>							
<b>IQ</b>	0.697	0.624	0.584	<b>0.810</b>						
<b>OI</b>	0.671	0.652	0.820	0.607	<b>0.891</b>					
<b>PEX</b>	0.722	0.661	0.759	0.630	0.719	<b>0.902</b>				
<b>SEQ</b>	0.622	0.587	0.626	0.557	0.574	0.570	<b>0.917</b>			
<b>SQ</b>	0.669	0.589	0.609	0.681	0.597	0.535	0.625	<b>0.834</b>		
<b>SU</b>	0.654	0.591	0.720	0.572	0.695	0.652	0.579	0.606	<b>0.869</b>	
<b>US</b>	0.768	0.674	0.813	0.701	0.778	0.742	0.666	0.714	0.685	<b>0.910</b>

APPENDIX D: VARIANCE INFLATION FACTOR (VIF)

	EEX	FC	II	IQ	OI	PEX	SEQ	SQ	SU	US
EEX									3.680	
FC									2.601	
II										
IQ									2.535	1.969
OI										
PEX									2.735	
SEQ									2.070	1.732
SQ									2.601	2.228
SU			1.882		1.882					
US			1.882		1.882				3.839	

# Autism Spectrum Disorder Detection: Video Games based Facial Expression Diagnosis using Deep Learning

Morched Derbali<sup>1</sup>, Mutasem Jarrah<sup>2</sup>, and Princy Randhawa<sup>3</sup>

Department of IT, Faculty of Computing and IT, King Abdulaziz University, Jeddah, Saudi<sup>1,2</sup>  
Department of Mechatronics Engineering, Manipal University Jaipur, Jaipur, India<sup>3</sup>

**Abstract**—In this study, a novel method is proposed for determining whether a child between the ages of 3 and 10 has autism spectrum disorder. Video games have the ability to immerse a child in an intense and immersive environment. With the expansion of the gaming industry over the past decade, the availability and customization of games for children has increased dramatically. When children play video games, they may display a variety of facial expressions and emotions. These facial expressions can aid in the diagnosis of autism. Footage of children playing a game may yield a wealth of information regarding behavioral patterns, especially autistic behavior. You can submit any video of a child playing a game to the interface, which is powered by the algorithm presented in this work. We utilized a dataset of 2,536 facial images of autistic and typically developing children for this purpose. The accuracy and loss function are presented to examine the 92.3% accurate prediction outcomes generated by the CNN model and deep learning.

**Keywords**—Autism in children; machine learning; deep learning; convolution neural network (CNN); video games; prediction

## I. INTRODUCTION

Autism is a complicated, behaviorally defined, static condition of an immature brain which is of significant concern to practicing pediatricians due to a staggering 55.6 per-cent rise in pediatric incidence from 1991 and 1997, surpassing spina bifida, cancer, and Down syndrome [1]. Rather than new environmental effects, this increase is due to increased awareness and evolving diagnostic criteria. Autism is a condition with numerous nongenetic and genetic origins, rather than a disease. Autism (autism spectrum disorders) is defined as a group of developmental disorders characterized by deficiencies in three behavioral domains: [2]

- 1) interpersonal interaction.
- 2) a diverse set of areas of interest and hobbies; and
- 3) speech, communication, and creative play.

### A. Autism and its Characteristics:

Early childhood autism is a pervasive developmental disorder. Autism affects communication, relationships, and self-control. Infants often get autism. Autism is a "spectrum disorder" that affects people differently. It has many traits. [3].

Early diagnosis can help a person with autism live a full life. According to the DSM-5, autism is characterized by

persistent differences in communication, interpersonal relationships, and social engagement [4]. Example: Being nonverbal or having abnormal speech patterns, having trouble understanding nonverbal communication, developing and maintaining relationships, and having trouble maintaining a traditional back-and-forth conversational manner [5]. Repetitive habits, interests, and behaviors. Excessive awareness to or significantly reduced sensitivity to many sensory stimuli, repetitive sounds or phrases (echolalia), preference for homogeneity and complexity with transition or regimen, rigid or heavily restricted and strenuous interests, hyper sensitivity to or dramatically reduced sensitivity to many sensory stimuli, rigid or severely regulated and intense interests, hyper sensitive. According to the American Psychological Association's Diagnostic statistical, autistic traits must be present in early childhood, but they may not fully express until social pressure builds the person's strength to deal with them, and difficulties may be covered up by learned coping skills [6].

### B. The Role of Video Games in Autism

According to research, enabling youngsters to play games using smart phones might help detect autism. Dr. Jonathan Delafield-Butt, a senior professor in childhood development, said it was important to detect autism early so parents and children could receive a variety of support services. [7]. Autism is a neurodevelopmental disorder with many shared traits, challenges, and abilities. Many autistic people have visual-spatial thinking, pattern identification, and a visual preference. Games that require visual clues and spatial skills are rewarding to such people. Games are creative but structured. RPGs and scrolling shooters satisfy research participants' desire for imagination without requiring self-generated creativity, which many autistic people lack. Video games have many audio and visual cues. Autistic people value rules and objectivity more than neurotypical people. To avoid anxiety and sensory meltdowns, follow clear guidelines. Video games reinforce clear expectations. Autistics need routine and repetition. Unknown circumstances cause anxiety, discomfort, and a desire to escape. Video games allow for controlled practice and mastery [8,9]. Games are more controlled than real life. Autism makes unpredictable human behavior difficult. Understanding social signs, idioms, humor, sarcasm, and satire can also cause anxiety. Playing a game that becomes more familiar each time helps autistic gamers overcome these challenges in a safe, controlled environment.

Parents and educators worry that autistic students spend too much time gaming instead of socializing. Playing has many benefits if encouraged and controlled [10].

## II. RELATED WORK

Modern diagnostic tools for mental diseases were developed in the late 1800s, although their origins may be dated back to the 4th century B.C (Before Christ) Era [11]. The gold standard for diagnosing often these mental-disorders relies heavily on information gathered from various respondents (e.g., parents, teachers) about the onset, direction, and duration of various behavioral descriptors, which is then considered by providers when making a diagnosis predicated on DSM-5 (Diagnostic and Statistical Manual) Categorization of Diseases-10th Installment (ICD-10) requirements [12]. Providers employ a variety of strategies to collect this data, ranging from subjective (e.g., assessment scale) and unstructured (e.g., semi - structured or unstructured interviews) to much more objective (e.g., actual observations) and organized (e.g., structured diagnostic interviews) [13].

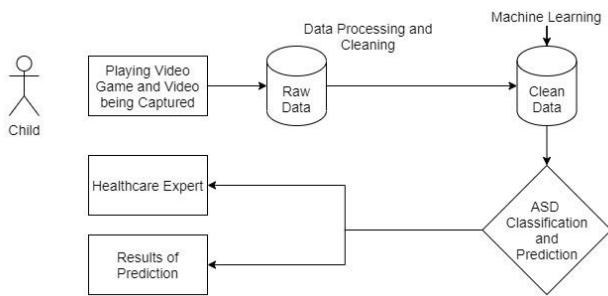


Fig. 1. Machine learning model for prediction and classification

Fig. 1 shows the architecture of the machine learning model that is used commonly for prediction of machine learning along with video games. The video of the child playing the game is captured and then the data is preprocessed, and machine learning is applied on it to predict the results.

Autism spectrum disorder (also referred to as ASD) and attention deficit hyperactivity disorder (also referred to as ADHD) are both conditions that are relatively common in children and can continue into adulthood. Autism spectrum disorder (ASD) is a developmental condition that causes patients to have difficulties with speech, behavior, and social interaction. Patients with ASD also tend to engage in repetitive behavior, have impatience issues, and attention problems. Since the publication of the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), the term autism spectrum disorder (ASD) has been used to refer to a more comprehensive diagnostic entity that formerly referred to a number of distinct disorders, including Autistic Disorder, Asperger's, and other Pervasive Chromosomal Anomalies [14]. According to findings from recent studies, the prevalence of autism spectrum disorder (ASD) in children and adolescents has increased from 1 in 100 to 1 in 59 in the past 14 years (from the year 2000 to 2014). Inattention, hyperactivity, and impulsivity are all symptoms of attention deficit hyperactivity disorder (ADHD), which is a common

brain condition found in children and adolescents. Both autism spectrum disorder and attention deficit hyperactivity disorder (ADHD) tend to continue into adulthood. A diagnosis of the condition is made in approximately 5.9–9.4% of all infants. Because autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) are so common in children, accurate and timely diagnosis of these conditions is critical. [15,16,17].

The field of artificial intelligence known as machine learning has the potential to significantly improve the use of computer methods in the field of neuroscience. A significant amount of research has been done to establish machine-learning models and deep-learning approaches in order to interpret high-dimensional MRI (Magnetic Resonance Imaging) data in order to simulate neural networks that regulate the brains of people who suffer from a variety of mental illnesses. [18,19]. These studies resulted in the development of machine-learning methods for the classification of Alzheimer's disease, mild cognitive impairment, right temporal epilepsy, schizophrenia, Parkinson's disease, dementia, attention deficit hyperactivity disorder, autism spectrum disorder, and major depressive disorder [20]. These statistical algorithm-based machine-learning models are well-suited to complex issues that require a combinatorial explosion of options or non-linear processes. This is because typical computer models fail in terms of quality or scalability in these situations, but these statistical algorithm-based machine-learning models succeed [21].

Senju et al, discussed the approaches to early detection of autism in infants. Early detection here refers to before 18 months of age. It gives an overview of the known processes of early social development in children that can fall under the category in which "core deficits" are manifested in young children and summarizes a criterion for the same. The paper concludes by discussing how the preferences for social stimuli changes in the infants over time [22]. They develop a preference for familiar people in first couple of months of life. They begin to have one to one interaction with their caregivers between three and six months of age. This is the period where they understand the interactive styles of their usual social partners. During four to seven months, infants can differentiate between emotional expressions of the face. Soon they develop preferences for caregivers and after 12 months, they display a pattern of response to separation and reunion that demonstrates what they have learned to expect from their experiences with their caregivers [23].

So, in the case of the infants with autism, one would speculate that the dyadic interactions would be less intense and lower capacity to recognize the emotional expressions of others. However, by three to five years of age, autistic children would be capable of forming secure attachments to their caregivers. The author concluded that similar patterns of attachment security and insecurity are found in typically developing children, it is very unlikely to be considered as a core deficit in autistic infants. In infants, failure to discriminate emotions can be considered as an important feature to characterize autism. But however, the rate of false positives could be very high. Children with autistic tendencies make use of different methods and skills to develop secure

attachments and to obtain the capacity to differentiate between the facial expressions and emotions than the typically developing children.

Whalen et al. [23,24] discovered that using a computer game to teach children with autism improved motivation and engagement when compared to the traditional methods of trying to teach children with autism. They established the computer game Teach Town to assist autistic children improve social, emotional, academic, and adaptive skills [25]. Whalen's findings are in line with what others in the area have found: video-game-like tools are beneficial with students with ASD because they are constant and predictable, entail few social variables, and enable children to control and set the speed of the activity. Many children on the autism spectrum have visual-spatial abilities that create videos actual gameplay an area of expertise. While it's normal to be concerned about inattention, behavioral concerns, and even addictions, there really are simple ways that may help autistic children get the most out of computer games and computers without causing additional problems at home or at school. The digital world provides a wealth of resources for teaching issue, social skills, adaptability in new contexts, and even motor skill development. Raising autistic children may be difficult, but apps, games, and technologies can make contacting and educating autistic children simpler. Tbatrah et al [26] laid out a principle called anger superiority hypothesis as its foundation. This hypothesis states that "angry faces capture attention faster than happy faces" in common individuals. The authors aim to test and compare the threat detection abilities in autistic people with individuals with a history of typical development using a facial visual search paradigm.

Thabtah et al [26] theory is based on building classification systems using machine learning, specifically a new method called Rules-machine Learning. This approach helps in detecting autistic traits and offers user knowledge bases (rules) which enables the professionals to make better analysis of the reasons behind the classification.

The primary objective of this technique is rule discovery by search method which can be done using covering classification. Then, evaluation is performed on the discovered rules to discard any redundancies and to optimize it further by reducing the number of discovered rules. In order to improve the overall efficiency of the training process, this phase contributes to the narrowing of the search space for individual data items. The classifier, which is utilized to make predictions regarding the value of the class, is more comparable to an outcome of the rule evaluation phase described above. For the purpose of generating the necessary data from the participants, a mobile application known as ASD Tests is utilized [27]. It implements four screening methods for toddlers, children, adolescents, and adults based on the Q-CHAT-10 (Quantitative Checklist for Autism in Toddlers), AQ-10-child, AQ-10-Adolescent, and AQ-10 adult, respectively. The Q-CHAT-10 is a quantitative checklist for autism in toddlers. In addition to this, the author made use of the datasets that were previously deposited in the University of California Irvine Data Repository by the authors.

Wu et al [28] analyzed the performances of various ML techniques such as Bagging, Boosting, rule induction, and decision tree classifiers on child, adolescent and adult ASD screening datasets. The error rates of adult dataset for the above techniques were between 5.68 and 8.23 per cent whereas the Rules Machine Learning (RML) model outperformed them with an error rate less than 5.6 per cent. Hence the paper was concluded by clearly revealing how ML approaches like covering can be used for obtaining promising results [29].

Jacob et al [29] obtained a high standard clinical data of children at risk for ASD to implement machine learning algorithms. The aim is to build a low-cost and easy to use ASD screening tool. To implement this, the author chose to proceed with a combination of two approaches. Two different algorithms are trained to combine their outputs as a final screening assessment. One is based on the short, structured parent- reported questionnaires and the second is based on tagging key behaviors from casual home videos of the test subjects. The first classifier was trained using data from ADI-R (Autism Diagnostic Interview) score sheets with labels corresponding to established clinical diagnoses. The training of second classifier i.e., the video classifier was done using ADOS (Autism Diagnostic Observation Schedule) instrument score sheets and diagnostic labels. To ensure sufficient training volume, progressive sampling was used in both the cases. After evaluating multiple machine learning algorithms, the author chose Random Forests for its robustness against overfitting [30].

In the clinical sample, the results showed that the parent questionnaire classification approach performed better than some of the more established screening tools, such as the M-CHAT (Modified Checklist for Autism in Toddlers) and the CBCL (Child Behavior Checklist). By combining the two different methods of classification into a single evaluation, performance was improved. The author concluded by stating how ML can play a crucial role in enhancing the performance of the behavioral health screeners and how this research demonstrated a significant improvement over established screening tools for autism. The author also mentioned how the research demonstrated how this research demonstrated significant improvement over established screening tools for autism. Chorianopoulou et al. [31] presented an ML-based approach to early diagnosis of ASD from videos of infants by identifying specific behaviours from them. This approach was based on using videos of the infants. They used a dataset that contained 2000 short videos with various behaviours of interest, such as directed gaze towards faces or objects of interest, positive affect vocalization, and other similar behaviours, all of which were manually coded by expert raters [32]. This dataset was used to conduct their research.

The authors addressed the issue by employing a deep learning model that was image-based and that was based on facial behavior features. Gorriz et al [32] has applied the various feature transformation techniques such as Log, Z-score, sine functions to the collected datasets of toddlers, children, adolescents, and adults [33]. In the next stage, various classification techniques were implemented with these

transformed ASD datasets, to evaluate and assess their performance [33].

For toddler dataset, the median highest result was calculated by Adaboost for Log transformation, Adaboost and SVM (Support Vector Machine) for Scale transformation respectively as 99.06%. The mean highest result which was reported to be 98.77% was calculated by SVM for Log and Sine transformations. The maximum highest result was recorded to be 100% was calculated by Adaboost, GLMboost and SVM for all feature transformation methods and C5.0 for Scale transformation [27, 34].

For child dataset, the median highest accuracy of 100% was achieved by LDA (Linear Discriminant analysis) and PCA (Principal Component Analysis) for Log and Scale

feature transformations. The mean highest accuracy of 97.2% was achieved by Adaboost for Log and Scale, respectively. Finally, the maximum highest accuracy of 100% was achieved by all classifiers and feature transformation methods [34,35].

For adolescent dataset, the median highest accuracy of 95% was obtained by C5.0, LDA, PCA where LDA and PCA for both Log and Scale as well as C5.0 for Scale. The mean highest result of 93.89% was obtained by PCA for Log and GLMboost for Scale, respectively. The maximum highest result of 100% were achieved by all classifiers and feature transformation methods [35,36].

Table I summarizes previous studies and its findings and limitations on usage of algorithms and the different features for the autism spectrum disorder

TABLE I. SUMMARIZED PREVIOUS STUDIES AND ITS FINDINGS FOR AUTISM SPECTRUM DISORDER

Reference	Findings	Limitations
Thabtah et al, 2018 [6]	Inside the ASD screening tool, SVM was used to integrate the ML algorithm. 97.6 % accuracy	Datasets were not balanced. The sample is small comprising 612 autism patients and 11 non-autism instances.
Thabtah et al, 2020 [26]	Rules-Machine Learning is a machine learning approach based on rule induction (RML). Covering learning was used to generate non-redundant rules in a simple method. RML classifies with greater prediction accuracy than typical algorithms such as boosting, bagging, and decision trees, thanks to the use of ten times cross-validation to split the dataset into ten subsets.	In terms of class labels, RML appeared to be ineffective when dealing with unbalanced data sets. There were no examples of toddlers in this article.
Vaishali et al, 2018 [27]	Optimal feature selection was automated using the Binary Firefly algorithm (ten out of twenty-one features were chosen as the best). There was no concern with class imbalance (there are 151 occurrences with class 'yes' and 141 instances with class 'no' in the ASD youngster's dataset). Models such as NB, J48, SVM, and KNN were used. SVM obtained the highest accuracy of 97.95 percent.	In the ASD kid dataset, there were some occurrences that are missing. There was a risk of model overfitting on the dataset because to the smaller number of occurrences in the dataset. Swarm intelligence wrappers had certain drawbacks (Binary Firefly algorithm)
Al banna et al, 2020 [33]	Analyzed the patient's condition using facial expressions and emotions, employing an AI system and sensor data. Sent out frequent messages to parents, assisting the patient in coping with ASD during COVID-19. A smart wristband with an integrated monitor and camera is linked to a smartphone app in this system. Used real-time grayscale photos from one Kaggle dataset of 35,887 images to detect ASD. The Inception-ResNetV2 architecture had the greatest accuracy of all the models, at 78.56 percent.	When compared to other methods, the accuracy is poor. The research is still in its early phases.
Sen et al, 2018 [37]	The authors devised a new algorithm that combined structural and functional characteristics. Drew many different depictions of the brain's functional connections. The results showed that incorporating multimodal characteristics improves case discrimination accuracy the most.	In contrast to earlier studies, the ML models utilized demonstrate a 4.2 percent improvement in the accuracy of the predictions for Autism. Datasets suffer significantly from fluctuations.
Van den et al, 2017 [39]	SVM, Naive Bayes, and Random Forest classification algorithms are used. There were 95,577 kid records with 367 variables, of which 256 were deemed to be adequate. Different qualities were well delineated. Created a dataset with four classifications (ASD: None, Mild, Moderate, Severe). The J48 algorithm attained the highest accuracy of 87.1 percent (2 class) and 54.1 percent (4 class) (decision tree)	Doesn't predict the severity of ASD. A cursory collection of traits (criteria) used to identify ASD, which may or may not always correspond to an instance of ASD.

### III. METHODOLOGY

This study demonstrated the use of Deep Learning and Image processing techniques for the detection of Autism using facial expressions. The initial approach was to build and train a neural network based on the available data on Autism. Following this, any video of the patient which clearly shows their facial expressions could be taken as the input through an interface created for the users. This input was used for the detection of autistic characteristics using the previously trained model [38].

The methodology followed can be divided into five steps as shown in Fig. 2:

- 1) Capturing the facial expressions while playing a Video Game.
- 2) Data Preprocessing.
- 3) Model Building and Training.
- 4) Prediction and Optimization.
- 5) Uploading the video to the Web Interface.

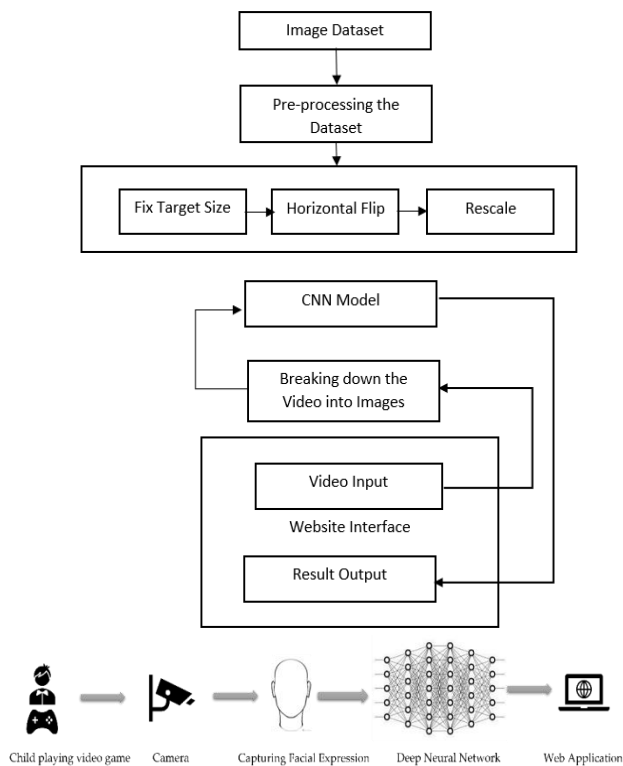


Fig. 2. Work flow of the method for the detection of facial expression while playing video games

#### A. Capturing the Facial Expressions while Playing the Video Game:

The facial expression of the child is captured while she/he is playing the video game using the web camera attached to the personal computer. The video is captured and saved locally and then the same video is uploaded to the website user interface and then the trained Convolutional Neural Network (CNN) model predicts if the child in the image broken down from the video is autistic or not [40].

#### B. Data Preprocessing

The collected image dataset was preprocessed by 3 processes so that the CNN model can train the dataset to predict if the child is autistic or not. They were:

1) *Dividing the video into frames:* OpenCV was used to fragment the video captured into images or frames. The video is captured using a webcam while the child is playing the game to monitor and analyze the video captured which would be broken down into images and then train the CNN model. The dataset consists of 2536 images belonging to the autistic and non-autistic classes in the training set and 300 images in the test set. The required image data is obtained from Kaggle. The ImageDataGenerator class of keras library enables us to read the images from the folders. Rescaling of images is done by dividing each pixel value of an image by 255. The images in the dataset are of various sizes which calls for resizing into one final size of 64x64 pixel and the same can be implemented using a function called “flow\_from\_directory” of the above class [40,41].

2) *Fix target size:* The collected image dataset had to be resized for uniformity in processing and to do so, all the images were resized to 64x64 pixel so that the convolution neural network model can be trained using those images [42]. This was done by using the function “flow\_from\_directory” and the argument to the function changed to 64x64 pixels. Whereas the neural network was training on the training data, the flow from directory () function was used to read photos straight from the directory and enhance them. The technique assumed that photos from various classes stored in separate directories but are all included within the same parent directory.

3) *Horizontal flip:* The images needed to be flipped horizontally to maintain uniformity, so this is achieved by using the function “ImageDataGenerator” and adjusting the arguments.

4) *Rescaling:* The ImageDataGenerator class may rescale pixel values from 0-255 to the recommended 0-1 range for neural network models. Normalization is the process of reducing data to a number between 0 and 1. Setting the rescale parameter to a ratio that may be multiplied by each pixel to produce the required range will do this [41,42].

#### C. Model Building and Training:

Deep learning is a technique for automatic learning that implements the use of examples to teach machines how to learn in the same way that people do. A self-driving car's ability to recognize a stop sign or differentiate between a pedestrian and a lamppost is dependent on its use of deep learning, which is an essential component of the technology. It makes it possible to control consumer electronics with one's voice, such as mobile phones and tablets, televisions, and hands-free speakers. The concept of "deep learning" has been receiving a lot of attention as of late, and with good reason. It's about achieving things that weren't possible before you started working on them.

During the process of deep learning, a computer model will learn to perform categorization tasks directly from either pictures, text, or sound. Models that use deep learning have the potential to achieve an accuracy that is on par with or even exceeds that of humans in certain circumstances. In order to train models, a significant amount of labelled data as well as various topologies of multilayer neural networks are utilized. The acronym "CNN" stands for "Convolution Neural Network". Image recognition and processing are two applications that make use of a type of artificial neural network known as a convolutional neural network (also abbreviated as CNN). CNNs are designed to focus specifically on analyzing pixel input. CNN was used to train the model based on the image data sets that were taken by the webcam of the computer that the child was using to play the video game on. CNNs are image processing, artificially intelligent (AI) systems that utilize deep learning to perform both generating and informative tasks. These tasks frequently include machine vision, which includes image and video identification, as well as recommendation systems for natural language processing (NLP) [40, -42]. A neural network is a piece of computer hardware and/or software that mimics the way neurons in the

human brain communicate with one another. Traditional neural networks were not intended to be used for image analysis, so in order for them to do so, they require the images to be broken up into smaller chunks. The "neurons" that make up CNN are organized more similarly to those in the prefrontal cortex, which is the part of the brain in humans and animals that is responsible for processing visual input. The difficulty of processing images in pieces that is inherent to traditional neural networks can be circumvented by arranging the layers of neurons in such a way that they cover the entirety of the visual field [48]. A CNN makes use of a technology similar to a perceptron that is designed to have minimal requirements for processing [43]. The layers of a CNN are comprised of an input layer, a layer, and a hidden layer. These layers are followed by several convolutional layers, average pooling, fully connected layers, and normalizing layers. A system that is significantly more effective and easier to train for image analysis and natural language [43] has been made possible as a result of the elimination of constraints and improvements in the efficiency of image processing.

#### D. Prediction and Optimization

Optimization plays a crucial role for any machine learning problem. Gradient descent is an optimization algorithm that finds the lowest possible value or the minimum value of a function through iterations. While the loss function which is also known as the cost function is all about calculating the loss/errors for every prediction that the neural network makes, gradient descent can be used to find the minimum of this loss function. The goal is to estimate the values of coefficients of a function that can minimize the cost function. In other words, the new coefficients will have a significantly lower cost.

This technique is initialized by taking small random values as coefficients for the function. The cost is then evaluated by inserting them into the function. The next step is to change the values of the coefficients in a direction that can lead to a lower cost in the next iteration. This direction can be estimated with the help of derivatives. The derivative gives a slope (gradient) at the desired point on a curve, or a function and the sign of that slope can be used to determine the direction in which the coefficients can be moved in further iterations [44]. Now that the algorithm is aware of the direction of progression of coefficients with the help of the gradient at current position, the next move is to make a step by scaling it and subtracting the obtained value from the current position. Subtracting is done as the aim is to minimize the function [40]. Another parameter called learning rate is used to scale the gradient and control the step size. Learning rate can affect the performance in a significant way. Smaller learning rate can lead the algorithm to reach the final iteration before even reaching the optimum point [41]. The Adam algorithm is implemented by

this optimizer. Adam optimization is a gradient descent approach based on adaptive first- and second-order moment estimation. The approach is "computationally more efficient, has small memory demand, is robust to diagonally resizing of gradient, and is well suited for situations with huge data/parameters," as according to Thabtah- et al., 2020.

#### E. Uploading the video to the Web Interface

Streamlit is one of the recent and fastest python-based model deployment tools. This open-sourced python based framework simplified the whole model deployment cycle along with providing an easy way to structure the functionalities of the interface [45,46].

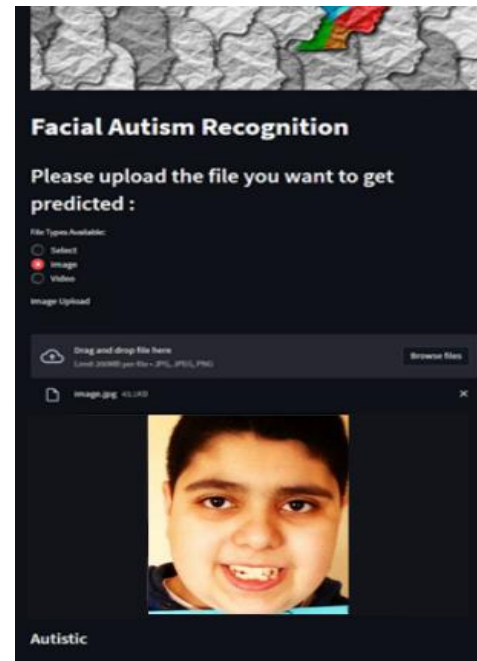


Fig. 3. Web interface for prediction of the disease

Once the neural network was trained, the model of an epoch with the best validation accuracy was saved as a '.model' file. The interface enables the users to select the type of media file that is to be uploaded. If the selected media type is an image, the saved model is used to classify the uploaded image to Autistic/Non-Autistic for which the result can be displayed on the interface [45]. But if the selected media type is a video, then the uploaded video is divided into frames using OpenCV and each image is classified into Autistic/Non-autistic with the help of the saved model. The mean value of classification of all images is considered as the final classification for the entire video and displayed on the interface.



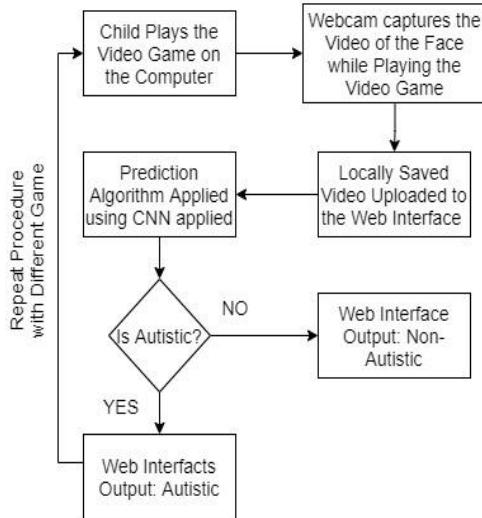


Fig. 4. Prediction mechanism if autistic or not

Fig. 3 describes the entire working of the prediction mechanism implemented in this project. The video of the child while playing the game is captured by a webcam and then the video is saved locally. The same video is then uploaded to the website interface for the CNN deep learning algorithm to work on it. After processing the file and after the prediction model is applied on the video, the outcome or result of the prediction is shown on the interface as “Autistic” or “Non-Autistic” as shown in Fig. 4.

#### IV. RESULTS AND DISCUSSION

This study explores the use of video games to discriminate between children with and without ASD. Compared to previous qualitative techniques, the activities and learning metrics in the evaluation game give a quantifiable depiction of children’s abilities, making the identification of ASD more accurate and practical. The use of these games as a supplementary tool in educational interventions for kids with ASD is also possible. In previous studies, the model has been trained using different algorithms such as SVM, Neural Network, RML Classifier, Random Forest, Naïve Bayes with a maximum accuracy of 97.5% [46]. They have used images of a facial expression and behavior of a patient while in our study we trained the model using CNN when the child is playing video game and capture the image of his behavior and facial expression. The Convolutional Neural Network (CNN) model is used to extract and produce the proper patterns of the face features when the child is playing video game and capture the image of his behavior and facial expression such as inappropriate snickering and laughing, Lack of pain sensitivity, Inability to maintain proper eye contact, unable to communicate with gestures, Inadequate reaction to sound etc. The model was trained in the cloud using Google Colab with python, which supports TensorFlow and Keras. The epoch number declared was 159 and it had a batch size of 20. The VGG model was used for implementing the convolutional neural network. Karen Simonyan & Andrew Zisserman of Oxford University’s Visual Geometry Group (VGG) proposed VGG models, which performed well in the ImageNet

Challenge [47 -48]. This model gets 92.7% top-5 accuracy on ImageNet’s 14 million photos from 1000 classes. 300 images were used for the testing and 100 images were used for validation of the CNN model [48].

The model achieved 92.3% accuracy for the testing dataset and 87.3% accuracy for the validation dataset. For this study, many evaluation metrics were calculated, and the results are summarized in the following sections. The results of the comparative prediction analysis of our model and existing model of the same dataset as shown in Table II.

TABLE II. COMPARISON OF THE RESULTS USING THE SAME DATASET

Author	Model	Accuracy
Haque et al [49]	ResNet50	89.2%
Shaik et al [50]	VGG19	84.0%
Our model	VGGFace	92.3%

#### A. Evaluation Metrics

Sensitivity, Specificity, Precision, Recall, F1 Score and Accuracy were calculated for evaluating the performance of the CNN model. The convention True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) were used.

1) *Sensitivity*: Sensitivity measures a model’s ability to recognize positive cases. TPR is sometimes called recall. Sensitivity measures how many positive cases a model accurately identified. High-sensitivity models have minimal false negatives, meaning they overlook some positive examples. Sensitivity is a model’s capacity to recognize good cases. We need our models to locate all good examples to generate accurate forecasts. True positive rate plus false negative rate equals 1. Higher true positive rate means model correctly identifies positive situations. The sensitivity was calculated using the following formula as shown in eq. (1):

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (1)$$

The Sensitivity of the model was 0.9560 or 95.60%.

2) *Specificity*: Specificity assesses the model’s ability to identify real negatives. This means there will be a percentage of true negatives forecasted as positives, or false positives. True Negative Rate (TNR). True negative rate plus false positive rate equals 1 always. Low specificity suggests the model is mislabeling a lot of negative data as positive. The specificity was calculated using the following formula as shown in eq. (2):

$$\text{Specificity} = \frac{TN}{FP+TN} \quad (2)$$

The specificity of the model was 0.8865 or 88.65%.

3) *Precision*: Precision is the ratio of True Positives to total positive samples (either correctly or incorrectly). It is calculated using the formula as shown in eq. (3):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

The precision of the model was 0.9048 or 90.48%.

4) *Accuracy*: Model accuracy measures which model is better at finding correlations and patterns in a dataset based on training data. Accuracy is calculated using the formula as shown in eq. (4):

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (4)$$

The accuracy was found out to be 92.3% for the testing dataset consisting of 300 images and 87.3% for the validation dataset.

5) *F1 score*: The harmonic mean of accuracy and recall is used to get the F1 score. It is calculated using the formula as shown in eq. (5):

$$\text{F1 Score} = 2 * \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

The F1 Score was calculated to be 0.9297 or 92.97%.

6) *Confusion matrix*: A matrix called the confusion matrix is used to assess how well classification models perform given

a particular set of test data. As the model cannot afford to predict non-autistic when the patient is autistic, the confusion matrix was calculated as shown in Fig. 5, which indicates that the false positive condition case value should be lower. 155 of the 300 images were correctly classified.

	Autistic	Non-Autistic
Predicted Autistic	152	16
Predicted Non-Autistic	7	125

Fig. 5. Confusion matrix using CNN for autism spectrum disorder

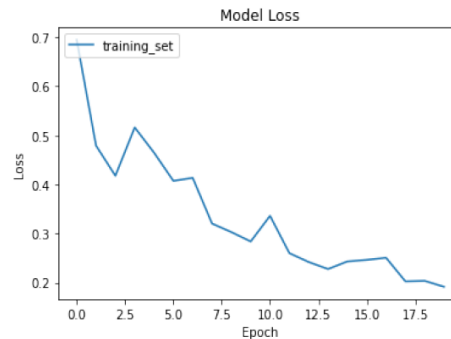


Fig. 6. a) and b) . Model Training Accuracy and Training Loss

Fig. 6(a) and (b) depicts the plot of the model's training accuracy, training loss, respectively for 10 epochs. The highest accuracy achieved was 92.3 % with the VGG model by applying two dense layers with specific parameters. A loss function optimizes ML algorithms. The loss is determined during training and validation, and its interpretation depends on how well the model performs. It's the total of training or validation set mistakes per example. Loss value indicates model's performance after each optimization cycle. A performance metric measures the algorithm's accuracy. Model accuracy is generally estimated as a percentage based on model parameters. It measures how well your model predicts actual data.

## V. CONCLUSION

The study built a deep learning web app to diagnose autism using a convolutional neural network and camera footage of a youngster playing a video game. CNN's architecture can extract facial attributes by generating facial feature patterns and assessing facial landmark distances, classifying faces as autistic or not. VGG CNN Model produced accurate results. Testing accuracy was 92.3%, validation accuracy was 87.3%, and precision was 90.4%. Future research will improve this model by broadening psychologists' autistic kid diagnoses. This programme helps

identify ASD. A precise autism diagnosis can help pick a treatment plan for autistic children. More precision might improve autism diagnosis. The platform might reveal this neurological disease, bringing treatment closer. This research is part of how humans utilize technology to tackle the world's healthcare issues. Future studies may use machine learning and deep learning algorithms to help individuals recognize a range of ailments using the same platform. Although young, digital technologies offer unlimited potential. In the case of autism, a mix of digital tools and in-person therapy visits is expected. Autism therapy isn't one-size-fits-all. These tools will help researchers produce medicines for ASD patients more swiftly. Language is being studied for potential therapies. SFARI-funded researchers are employing cellphones and automatic transcription software to record speech from autistic youngsters. Based on the data collected by recording footage of children playing video games, a video game can be designed for children with autism or autistic symptoms. This game will help parents or guardians determine whether their kid has autism such as Kinect game.

#### ACKNOWLEDGMENT

The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number 257835-G: 243-611-1443 and King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. This data can be found here: <https://www.kaggle.com/code/basmarg/vgg-16-autism-image-data/data>

**Conflicts of Interest:** The authors declare no conflict of interest.

#### REFERENCES

- [1] Aljero, M.K.A.; Dimililer, N. A Novel Stacked Ensemble for Hate Speech Recognition. *Applied Sciences* 2021, 11, doi:10.3390/app112411684.
- [2] Suman Raj, Sarfaraz Masood, Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques, *Procedia Computer Science*, Volume 167, 2020, Pages 994-1004, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2020.03.399>.
- [3] Chae, S.; Choi, A.; Jung, H.; Kim, T.H.; Kim, K.; Mun, J.H. Machine Learning Model to Estimate Net Joint Moments during Lifting Task Using Wearable Sensors: A Preliminary Study for Design of Exoskeleton Control System. *Applied Sciences* 2021, 11, doi:10.3390/app112411735.
- [4] Nishio, M. Special Issue on "Machine Learning/Deep Learning in Medical Image Processing". *Applied Sciences* 2021, 11, doi:10.3390/app112311483.
- [5] Pană, M.-A.; Busnatu, Ș.-S.; Serbanoiu, L.-I.; Vasilescu, E.; Popescu, N.; Andrei, C.; Sinescu, C.-J. Reducing the Heart Failure Burden in Romania by Predicting Congestive Heart Failure Using Artificial Intelligence: Proof of Concept. *Applied Sciences* 2021, 11, doi:10.3390/app112411728.
- [6] Thabtah, Fadi, Firuz Kamalov, and Khairan Rajab. (2018) "A new computational intelligence approach to detect autistic features for autism screening." *International journal of medical informatics* 117: 112-124.
- [7] Tamang, L.D.; Kim, B.W. Deep Learning Approaches to Colorectal Cancer Diagnosis: A Review. *Applied Sciences* 2021, 11, doi:10.3390/app112210982.
- [8] Abbas, H.; Garberson, F.; Glover, E.; Wall, D.P. Machine learning approach for early detection of autism by combining questionnaire and home video screening. *Journal of the American Medical Informatics Association : JAMIA* 2018, 25, 1000-1007, doi:10.1093/jamia/ocy039.
- [9] Akter, T.; Shahriare Satu, M.; Khan, M.I.; Ali, M.H.; Uddin, S.; Lio, P.; Quinn, J.M.W.; Moni, M.A. Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders. *IEEE Access* 2019, 7, 166509-166527, doi:10.1109/access.2019.2952609.
- [10] M. N. Parikh, H. Li, and L. He, "Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data," *Frontiers in Computational Neuroscience*, 2019.
- [11] Chu, H.-C.; Tsai, W.W.-J.; Liao, M.-J.; Chen, Y.-M. Facial emotion recognition with transition detection for students with high-functioning autism in adaptive e-learning. *Soft Computing* 2017, 22, 2973-2999, doi:10.1007/s00500-017-2549-z.
- [12] F. Thabtah and D. Peebles, "A new machine learning model based on induction of rules for autism detection," *Health Informatics Journal*, vol. 26, no. 1, pp. 264–286, 2020.
- [13] M. H. Al Banna, T. Ghosh, K. A. Taher, M. S. Kaiser, and M. Mahmud, "A monitoring system for patients of autism spectrum disorder using artificial intelligence," in *Proceedings of the International Conference on Brain Informatics*, pp. 251–262, Springer, Padua, Italy, July 2020.
- [14] Ganesh, K.; Umapathy, S.; Thanaraj Krishnan, P. Deep learning techniques for automated detection of autism spectrum disorder based on thermal imaging. *Proc Inst Mech Eng H* 2021, 235, 1113-1127, doi:10.1177/09544119211024778.
- [15] Deshpande, G., Wang, P., Rangaprakash, D., and Wilamowski, B. (2015). Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data. *IEEE Trans. Cybernet.* 45, 2668–2679. doi: 10.1109/TCYB.2014.2379621
- [16] A. V. Dahiya, C. McDonnell, E. DeLucia, and A. Scarpa, "A systematic review of remote telehealth assessments for early signs of autism spectrum disorder: video and mobile applications," *Practice Innovations*, vol. 5, no. 2, pp. 150–164, 2020.
- [17] Hassouneh, A.; Mutawa, A.M.; Murugappan, M. Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods. *Informatics in Medicine Unlocked* 2020, 20, doi:10.1016/j.imu.2020.100372.
- [18] Gwyn, T.; Roy, K.; Atay, M. Face Recognition Using Popular Deep Net Architectures: A Brief Comparative Study. *Future Internet* 2021, 13, 164
- [19] N. Zaman, J. Ferdus and A. Sattar, "Autism Spectrum Disorder Detection Using Machine Learning Approach," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 1-6, doi: 10.1109/ICCCNT51525.2021.9579522.
- [20] Elshoky, B. R. G., Younis, E. M. G., Ali, A. A., and Ibrahim, O. A. S., Comparing automated and non-automated machine learning for autism spectrum disorders classification using facial images, *ETRI Journal* 44 (2022), 613– 623. <https://doi.org/10.4218/etrij.2021-009721>.
- [21] Sato, W.; Sawada, R.; Uono, S.; Yoshimura, S.; Kochiyama, T.; Kubota, Y.; Sa-kihama, M.; Toichi, M. Impaired detection of happy facial expressions in autism. *Sci Rep* 2017, 7, 13340, doi:10.1038/s41598-017-11900-y.
- [22] Lee JH, Lee GW, Bong G, Yoo HJ, Kim HK. Deep-Learning-Based Detection of Infants with Autism Spectrum Disorder Using Auto-Encoder Feature Representation. *Sensors*. 2020; 20(23):6762. <https://doi.org/10.3390/s20236762>
- [23] Levy, S.; Duda, M.; Haber, N.; Wall, D.P.J.M.a. Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. 2017, 8, 1-17.
- [24] Król, M.E.; Król, M.J.N. A novel machine learning analysis of eye-tracking data reveals suboptimal visual information extraction from facial stimuli in individuals with autism. 2019, 129, 397-406.
- [25] Center for Autism Research. Social Responsiveness Scale, 2nd Edition (SRS-2). Available online: <https://www.carautismroadmap.org/social-responsiveness-scale/?print=pdf> (accessed on 26 October 2020).
- [26] Thabtah, F.; Peebles, D. A new machine learning model based on induction of rules for autism detection. *Health Informatics J* 2020, 26, 264-286, doi:10.1177/1460458218824711.
- [27] Vaishali, R., & Sasikala, R. (2018). A machine learning based approach to classify autism with optimum behaviour sets. *International Journal of Engineering & Technology*, 7(4), 18.
- [28] Wu, C.; Liaqat, S.; Helvaci, H.; Cheung, S.S.; Chuah, C.N.; Ozonoff, S.; Young, G. Machine Learning Based Autism Spectrum Disorder Detection from Videos. *Healthcom* 2021, 2020, doi:10.1109/healthcom49281.2021.9398924.
- [29] Jacob, S.; Wolff, J.J.; Steinbach, M.S.; Doyle, C.B.; Kumar, V.; Elison, J.T.J.T.p. Neurodevelopmental heterogeneity and computational approaches for understanding autism. 2019, 9, 1-12.
- [30] Chorianopoulou, A.; Tzinis, E.; Iosif, E.; Papoulidi, A.; Papailiou, C.; Potamianos, A. Engagement detection for children with Autism Spectrum Disorder. In *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5-9 March 2017, 2017; pp. 5055-5059.

- [31] Achenie, L.E.; Scarpa, A.; Factor, R.S.; Wang, T.; Robins, D.L.; McCrickard, D.S.J.J.o.d.; JDBP, b.p. A machine learning strategy for autism screening in toddlers. 2019, 40, 369.
- [32] Górriz, J.M.; Ramírez, J.; Segovia, F.; Martínez, F.J.; Lai, M.-C.; Lombardo, M.V.; Baron-Cohen, S.; Consortium, M.A.; Suckling, J.J.I.j.o.n.s. A machine learning approach to reveal the neurophenotypes of autisms. 2019, 29, 1850058.
- [33] Al Banna, M.H.; Ghosh, T.; Taher, K.A.; Kaiser, M.S.; Mahmud, M. A monitoring system for patients of autism spectrum disorder using artificial intelligence. In Proceedings of the International Conference on Brain Informatics, 2020; pp. 251-262.
- [34] Duda, M.; Ma, R.; Haber, N.; Wall, D.P.J.T.P. Use of machine learning for behavioral distinction of autism and ADHD. 2016, 6.
- [35] Sen, B.; Borle, N.C.; Greiner, R.; Brown, M.R.J.P.o. A general prediction model for the detection of ADHD and Autism using structural and functional MRI. 2018, 13, e0194856.
- [36] Thabtah, F.; Peebles, D.J.H.i.j. A new machine learning model based on induction of rules for autism detection. 2020, 26, 264-286.
- [37] Sen, B., Borle, N. C., Greiner, R., & Brown, M. R. (2018). A general prediction model for the detection of ADHD and Au-tism using structural and functional MRI. *PloS one*, 13(4), e0194856..
- [38] F. C. Tamilarasi and J. Shanmugam, "Convolutional Neural Network based Autism Classification," in 2020 5th Interna-tional Conference on Communication and Electronics Systems (ICCES), pp. 1208–1212, IEEE, (2020, June).
- [39] van den Bekerom, B. Using machine learning for detection of autism spectrum disorder. In Proceedings of the Proc. 20th Student Conf. IT, 2017; pp. 1-7.
- [40] Carette, R.; Elbattah, M.; Cilia, F.; Dequen, G.; Guérin, J.-L.; Bosche, J. Learning to Predict Autism Spectrum Disorder based on the Visual Patterns of Eye-tracking Scanpaths. In Proceedings of the HEALTHINF, 2019; pp. 103-112.
- [41] Guimarães, A.J.; Araujo, V.J.S.; Araujo, V.S.; Batista, L.O.; de Campos Souza, P.V. A hybrid model based on fuzzy rules to act on the diagnosed of autism in adults. In Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations, 2019; pp. 401-412.
- [42] Sumi, A.I.; Zohora, M.F.; Mahjabeen, M.; Faria, T.J.; Mahmud, M.; Kaiser, M.S. fassert: A fuzzy assistive system for chil-dren with autism using internet of things. In Proceedings of the International Conference on Brain Informatics, 2018; pp. 403-41
- [43] Hassan, T.-u.; Abbassi, R.; Jerbi, H.; Mehmood, K.; Tahir, M.F.; Cheema, K.M.; Elavarasan, R.M.; Ali, F.; Khan, I.A. A Novel Algorithm for MPPT of an Isolated PV System Using Push Pull Converter with Fuzzy Logic Controller. *Energies* 2020, 13, doi:10.3390/en13154007.
- [44] Tariq, Q.; Fleming, S.L.; Schwartz, J.N.; Dunlap, K.; Corbin, C.; Washington, P.; Kalantarian, H.; Khan, N.Z.; Darmstadt, G.L.; Wall, D.P.J.J.o.m.I.r. Detecting developmental delay and autism through machine learning models using home vide-os of Bangladeshi children: Development and validation study. 2019, 21, e13822.
- [45] Pavithra, D., Jayanthi, A.N., Nidhya, R. and Balamurugan, S. (2022). Autism Screening Tools With Machine Learning and Deep Learning Methods: A Review. In *Tele-Healthcare* (eds R. Nidhya, M. Kumar and S. Balamurugan). <https://doi.org/10.1002/9781119841937>.
- [46] P. Mazumdar, G. Arru, and F. Battisti, Early detection of children with autism spectrum disorder based on visual exploration of images, *Signal Process. Image Commun.* 94(2021), 116184.
- [47] Fawaz Waselallah Alsaade, Mohammed Saeed Alzahrani, "Classification and Detection of Autism Spectrum Disorder Based on Deep Learning Algorithms", *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 8709145, 10 pages, 2022. <https://doi.org/10.1155/2022/8709145>.
- [48] Miao, Y., Dong, H., Jaam, J. M. A., & Saddik, A. E. (2019). A deep learning system for recognizing facial expression in real-time. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 15(2), 1-20.
- [49] M. I. U. Haque and D. Valles, "A facial expression recognition approach using DCNN for autistic children to identify emotions," in Proceedings of the 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 546–551, Vancouver, Canada, 1–3 November 2018.
- [50] Jahanara, S., & Padmanabhan, S. (2021). Detecting autism from facial image, *International Journal of Advance Research, Ideas and Innovations in Technology*, Volume 7, Issue 2 - V7I2-1181.

# An Improved SVM Method for Movement Recognition of Lower Limbs by MIMU and sEMG

Xu Yun<sup>1</sup>, Xu Ling<sup>2</sup>, Gao Lei<sup>3</sup>, Liu Zhanhao<sup>4</sup>, Shen Bohan<sup>5</sup>

School of Information Science and Engineering, Zhejiang Sci-Tech University  
Hangzhou, China<sup>1, 2, 3, 4</sup>

School of Mechanical Engineering, Zhejiang Sci-Tech University  
Hangzhou, China<sup>5</sup>

**Abstract**—Aiming at the problems that the movement recognition accuracy of lower limbs needs to be improved, the optimized SVM recognition method by using voting mechanism is proposed in this paper. First, CS algorithm is applied to optimize the kernel function parameter and the penalty factor for SVM model. And then, voting mechanism is used to ensure the recognition accuracy of SVM classification algorithm. Finally, the experiments have been implemented and different classification algorithms have been compared. The recognition results shows that the movement recognition accuracy for the lower limbs by the optimized SVM recognition algorithm using voting mechanism is about 98.78%, which is higher than other commonly used classification algorithm with or without voting mechanism. The recognition method for the lower limbs proposed in this paper can be used in the field of rehabilitation training, smart healthcare and so on.

**Keywords**—Surface electromyography; micro inertial measurement unit; support vector machine; voting mechanism

## I. INTRODUCTION

Along with the development of artificial intelligence technology, many researchers have focused on the study of human posture recognition [1-4]. Recognition for the movements of lower limbs is widely used in the field of rehabilitation training, physical exercise and so on [5]. As for the current research, there are many sensors used for the recognition of lower limbs, such as visual sensors, inertial sensors, surface electromyographic sensors, etc.

In the study of human behavior recognition using visual sensors, Nie proposed a view-invariant method for human action recognition by recovering the corrupted skeletons based on a 3D bio-constrained skeleton model and visualizing those body-level motion features obtained during the recovery process with images [6]. Based on the Northwestern-UCLA dataset, the classification accuracy of the 10 action images in the proposed algorithm is about 94.40%. Nieto-Hidalgo proposed an extraction system by analyzing image sequences to identify human gait features [7]. The recognition accuracy for the normal and abnormal gaits are both more than 90.00%. Based on the Kinect sensor, Min proposed an indoor fall detection method using SVM method according to the 3D skeleton joint array information [8]. The experiment result shows that the fall recognition accuracy is about 92.05%. Liu proposed an algorithm for human behavior recognition using skeletal joint information of deep sequences [9]. The angle and

position information between joints were captured by RGB video, and the obtained feature vectors were used as the input of the classifier. The experimental results showed that the average accuracy of behavior recognition is 95.00%. It is convenient to recognize the human behavior by using visual sensors, but the environment illumination, image resolution, and autofocus speed will affect the recognition accuracy. What's more, the huge amount of image processing will increase the hardware cost.

In the study of human behavior recognition using inertial sensors, Khatun proposed a sensor-based learning method for human activity recognition [10] based on a hybrid deep learning model coupling convolutional neural network and long and short term memory network. The system is based on the data set (H-Activity) collected by the smart phone sensor. The experimental results show that the accuracy of the self-collected data set trained by this method is 99.93%, and the accuracy of the model trained by the benchmark data set (MHEALTH) is 98.76%. Zhang proposed a SVM algorithm based on magnetometer and gyroscope sensors to classify the human motion postures [11]. Experimental results show that the average recognition accuracy for human motion posture is about 90.80%. Marron proposed a smart phone system with embedded inertial sensors in an indoor environment [12], in which the information of human biomechanical models is combined. The recognition average accuracy for human behavior is 95.00%. Guo proposed a novel monitoring framework of human motion sequences based on wearable inertial sensors [13], the recognition process can be divided into data acquisition, segmentation and recognition stages. At the recognition stage the HMM algorithm is used to recognize the motion sequence. The experimental results show that the average recognition accuracy for human movement is 92.75%. As for the inertial sensors used for the recognition of human behavior, it has the advantages of high efficiency, but the installation will affect the recognition accuracy.

In the study of human behavior recognition using surface electromyography (sEMG) sensors, Qi proposed a gesture recognition system based on the principal component analysis method and GRNN neural network [14]. By extracting the key information of human gestures, the specific action mode can be identified. Experiment results show that the system's overall recognition accuracy for 9 static gestures is about 95.10%. Zhang proposed a dynamic adaptive neural network algorithm based on multi-feature fusion of surface EMG signals [15] to

achieve accurate recognition of eight lower limb movements (walk (WK), left turn (TL), right turn (TR), stand up (TP), sit down (ST), go upstairs (UPS), go downstairs (DWS) and jog (CD)). Experimental results show that the recognition accuracy of this method is 94.89%. As for the application of sEMG sensors, human behavior recognition can be easily affected by some irregular movements of the human body, such as when somebody suddenly falls, it may cause a significant change in the sEMG signal of the leg muscles, which will affect the judgment of human leg movements.

From the aforementioned methods for the recognition of human behavior recognition, the accuracy still needs to be improved. In this paper, take the features of lower limbs into consideration, the improved SVM algorithm is proposed for the recognition of lower limbs by MIMU and sEMG sensors. First, in order to obtain the parameters of kernel function and penalty factor, Cuckoo search algorithm (CS) is used to optimize SVM model. Then, the voting mechanism is applied to improve the recognition accuracy of SVM algorithm. Finally, the experiment is carried out to verify the validity of the proposed method by seven movements of lower limbs.

## II. CLASSIFICATION ALGORITHM

### A. SVM Classification Algorithm

As for the complexity of the movements of low limbs, it is difficult to recognize the movements directly from the outputs of MIMU and sEMG. Thus, it is necessary to extract the features of the outputs and then use a kind of mapping algorithm to separate the features from the low dimensional space to a high dimensional space. SVM classification algorithm based on the principle of structural risk minimization has the advantages of good generalization ability [16-18]. In this paper, according to the features of outputs of MIMU and sEMG and take the complexity of the algorithm calculation into consideration, SVM classification algorithm is used.

As for SVM classification algorithm, suppose the sample set is  $D=\{(x_1,y_1),(x_2,y_2),\dots,(x_m,y_m)\}$ , where,  $x_i \in R^n$ ,  $y_i \in \{-1,1\}$ ,  $m$  is the number of samples.  $\phi(x_i)$  is the eigenvector after mapping. Then, the hyperplane is established for classification by:

$$f(x) = W\phi(x_i) + b \quad (1)$$

Where,  $W$  is the normal vector of the optimal hyperplane,  $b$  is the displacement. The problem to solve the optimal hyperplane can be converted by:

$$\begin{cases} \min \left( \frac{1}{2} \|W\|^2 + c \sum_1^m \xi_i \right) \\ s.t. \quad y_i [W^T \phi(x_i) + b] \geq 1 - \xi_i \end{cases} \quad (2)$$

Where,  $\xi_i$  is the relaxation variable,  $c$  is the penalty factor. Introducing the Lagrange multiplier  $\alpha_i$  by:

$$\begin{cases} \max \left[ \sum_1^m \alpha_i - \frac{1}{2} \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j) \right] \\ s.t. \quad \sum_1^m \alpha_i y_i = 0, c \geq \alpha_i \geq 0, i = 1, 2, \dots, m \end{cases} \quad (3)$$

Where,  $\kappa(x_i, x_j)$  is the kernel function. And in this paper, the gaussian radial basis function is selected by:

$$\kappa(x_i, x_j) = \phi(x_i)^T \phi(x_j) = \exp(-g \|x_i - x_j\|^2) \quad (4)$$

Where,  $g$  is the kernel function parameter. Then, solving (3) and the following solution can be obtained:

$$f(x) = \text{sgn} \left( \sum_{i=1}^m \alpha_i^* y_i \kappa(x_i, x_j) + b^* \right) \quad (5)$$

In the process to solve the hyperplanes, the kernel function parameter  $g$  and the penalty factor  $c$  are the key parameters for SVM classification model. Thus, it is necessary to get the best kernel function parameter  $g$  and the penalty factor  $c$  for the SVM classification model to recognize the movements of the lower limbs.

### B. CS Algorithm

CS Algorithm is a heuristic algorithm, which solves the optimal parameters by simulating the parasitic brooding behavior of cuckoo birds [19,20]. It has the advantages of few parameters and fast convergence speed, so it is widely used for parameter optimization.

As for the parameter optimization of the kernel function parameter  $g$  and the penalty factor  $c$ , the finding probability  $p_a$  in CS algorithm will balance the local random optimization and global random optimization. The local random optimization will updates the position of the nest  $z=[g \ c]^T$  as follows:

$$z_i(n+1) = z_i(n) + \alpha s \otimes H(p_a - \varepsilon) \otimes (z_j(n) - z_i(n)) \quad (6)$$

Where,  $z_i(n+1)$  represents the position of the nest  $i$  updated at the  $(n+1)$  times of iteration,  $z_j(n)$  is the nest selected by random substitution,  $\alpha$  is the step scale factor,  $s$  is the step size,  $H(u)$  is the unit step function,  $\varepsilon$  is a random number.

Then, the global random optimization will update the position of the nest  $z=[g \ c]^T$  by:

$$z_i(n+1) = z_i(n) + \alpha L(s, \lambda) \quad (7)$$

Where,  $L(s, \lambda)$  is the Levy distribution.  $\lambda$  is a constant and in this paper it is 1.5 for the parameter optimization in CS algorithm.

As for the  $(n+1)$  times of iteration, the position of the nest will be updated, and it will be used as the input for the next iteration. According to the finding probability  $p_a$ , the highest accuracy nest  $z=[g \ c]^T$  for the whole nests will be obtained.

### C. Voting Mechanism

Voting mechanism proposed in this paper is a combination strategy for the movement recognition of lower limbs. The basic idea of it is using a sliding window to select the most recognition output label by the machine learning algorithm as the recognition output label for the middle position of this sliding window. Thus, it can correct the error recognition label.

The hardware for the recognition of the lower limbs in this paper is self-designed and the sampling frequency for the output data is 150Hz. According to the experiment for a consecutive movement of lower limbs, the self-designed hardware can collect at least 75 groups of output data. For the same movement recognition of lower limbs, it should be the same output label when using machine learning recognition algorithm. However, for the different movement recognition of lower limbs, the voting mechanism is necessary by the sliding window to identify and correct the error recognition label.

### III. EXPERIMENT FOR MOVEMENTS OF LOWER LIMBS

In order to verify the efficiency of the proposed optimize SVM classification algorithm, the movement recognition experiments for the lower limbs have been carried out. The designed recognition flow is shown in Fig. 1.

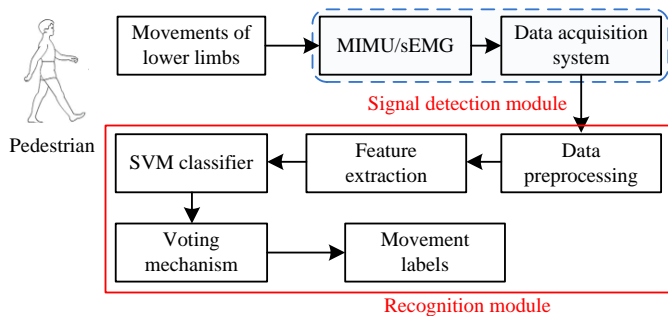


Fig. 1. Recognition flow for the lower limbs

The outputs of MIMU and sEMG will be collected by the signal detection module. And then the outputs will be preprocessed, during this process, the median filter is used to preprocess the outputs of the sensors which will remove the noise. Then, the features will be extracted by the mean absolute average (MAV) before they were sent to SVM classifier. What's more, the voting mechanism will be carried out to ensure the recognition accuracy. Finally, the movement labels for the movement of lower limbs will be shown.

#### A. Hardware for the Experiment

The self-designed hardware for the recognition of the lower limbs is shown in Fig. 2.

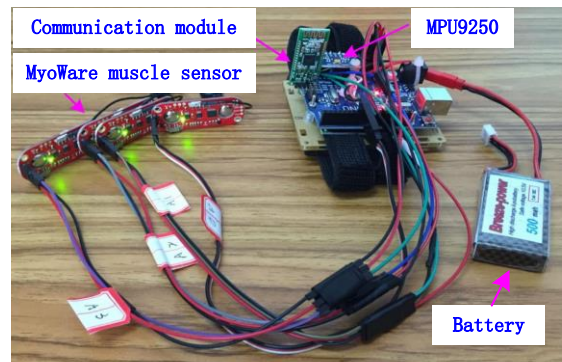


Fig. 2. Hardware for the experiment

MPU9250 and MyoWare muscle sensor is used for MIMU and sEMG. And the self-designed hardware will be worn on the lower limbs as shown in Fig. 3.

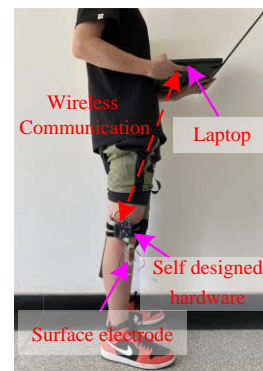


Fig. 3. Hardware worn on the lower limb

The muscles selected for the experiment should be close to the skin surface and there should be large enough to place the electrodes. Thus, we choose the tibialis anterior, extensor digitorum longus, gastrocnemius muscle, and soleus muscle to place the sEMG electrodes for the reason that the muscular contraction is obvious when there's movement of the lower limbs.

#### B. Definition of Movements of Lower Limbs

In this paper, there are seven movements of lower limbs needs to recognize, which are defined in Table I.

TABLE I. MOVEMENTS OF LOWER LIMBS

Movements of lower limbs	Labels
stand still	1
mark time	2
run with raised legs	3
go straight	4
run with speed	5
walk up the slope	6
walk down the slope	7

C. Recognition Test

There are 1072 sets data collect, we choose 872 sets of data as a fixed training set randomly, and then the following 200 sets of data are selected as the test set.

CS algorithm is used to optimize kernel function parameter  $g$  and the penalty factor  $c$  for SVM classification algorithm. As for CS algorithm, the number of parameters need to be optimized is 2, the number of nests is 20, the finding probability  $p_a$  is 0.25, the iteration time is 50, the upper bound of  $g$  and  $c$  is 10, and the lower bound is 0.01. Then, SVM classification model can be established.

As for the recognition, 30 independent recognition tests have been carried out based on 200 sets of data. Fig. 4 and Fig. 5 have illustrated one of the recognition results for the proposed optimized SVM classification model with and without voting mechanism.

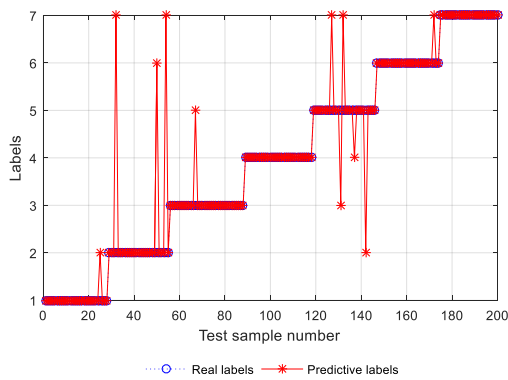


Fig. 4. SVM classification without voting mechanism

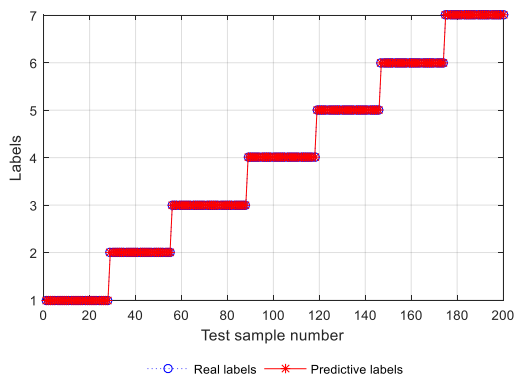


Fig. 5. SVM classification with voting mechanism

As show in Fig. 4 and Fig. 5, it is obvious that the recognition accuracy has been improved when voting mechanism is used to correct the wrong movements' labels of lower limbs labels.

In order to verify the effectiveness of the proposed optimized SVM recognition method in this paper, the most commonly used algorithm for human motion recognition has been applied, such as generalized regression neural network (GRNN), probabilistic neural network (PNN), and extreme learning machine (ELM). 30 independent recognition tests have been performed by using the above algorithms with and

without voting mechanism. The recognition accuracy is shown in Fig. 6 and Table II.

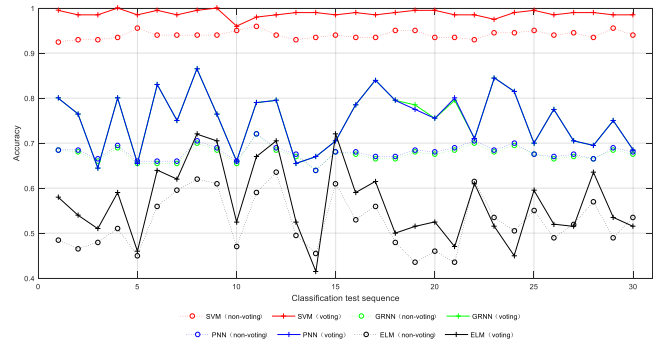


Fig. 6. Recognition accuracy for voting and non-voting mechanism

TABLE II. STATISTICAL RESULTS FOR THE RECOGNITION

Recognition Methods		Standard Deviation	Average accuracy	Coefficient of Variation
optimized SVM	Non-voting	$8.55 \times 10^{-3}$	94.05%	$9.09 \times 10^{-3}$
	voting	$7.73 \times 10^{-3}$	98.78%	$7.83 \times 10^{-3}$
GRNN	Non-voting	$1.66 \times 10^{-2}$	67.57%	$2.46 \times 10^{-2}$
	voting	$6.32 \times 10^{-2}$	75.28%	$8.40 \times 10^{-2}$
PNN	Non-voting	$1.66 \times 10^{-2}$	67.97%	$2.44 \times 10^{-2}$
	voting	$6.32 \times 10^{-2}$	75.27%	$8.40 \times 10^{-2}$
ELM	Non-voting	$6.03 \times 10^{-2}$	52.47%	$1.15 \times 10^{-1}$
	voting	$8.30 \times 10^{-2}$	56.77%	$1.46 \times 10^{-1}$

Fig. 6 and Table II show that the average recognition accuracy of lower limbs by the optimized SVM recognition method is higher than other commonly used algorithm for human motion. What's more, the recognition accuracy by the voting mechanism is higher than that without voting mechanism.

For the statistical results of 30 independent recognition tests show in Table II, the standard deviation and the coefficient of variation for the proposed optimized SVM recognition method with the voting mechanism are the minimum values in the above methods, which indicate the stability of the proposed method.

IV. CONCLUSION

Aiming at the problems that the recognition accuracy and stability for the movements of lower limbs need to be improved, the optimized SVM recognition method by using voting mechanism is proposed in this paper. CS algorithm is used to optimize kernel function parameter  $g$  and the penalty factor  $c$  for SVM classification algorithm. Then, voting mechanism is carried out to ensure the recognition accuracy. Experiments have been implemented and different classification algorithms have been compared. The recognition results illustrated that the recognition accuracy by the optimized SVM recognition algorithm is higher than other recognition algorithms, which demonstrates the validity of the proposed optimized SVM recognition algorithm. And the accuracy for the recognition algorithm with the voting



mechanism is higher than the algorithm without the voting mechanism, which demonstrates the effectivity of the voting mechanism. The recognition method for the movement of lower limbs by MIMU and sEMG can be used for the application of smart healthcare, rehabilitation training of lower limbs and so on.

#### ACKNOWLEDGMENT

This study was supported by Zhejiang Provincial Natural Science Foundation of China Grant Number LQ20F030019. And it was supported by the National Natural Science Foundation of China Grant number 62203393.

#### REFERENCES

- [1] Q. Liu, "Aerobics posture recognition based on neural network and sensors," *Neural Comput. Appl.*, vol. 34, pp. 3337-3348, January 2022.
- [2] C. Wang, F. Zhang, and X. Zhu, "Low-resolution human pose estimation," *Pattern Recogn.*, vol. 126, pp. 1-10, January 2022.
- [3] M. Panwar, D. Biswas, and H. Bajaj, "Rehab-net: deep learning framework for arm movement classification using wearable sensors for stroke rehabilitation," *IEEE Trans. Biomed. Eng.*, vol. 66, pp. 3026-3037, December 2019.
- [4] J. Huang, X. Zou, and Z. Fan, "Human pose estimation and LSTM-based diver heading prediction for AUV navigation guidance," *Signal Image Video Process.*, vol. 16, pp. 395-402, July 2022.
- [5] M. Markovic, M. Varel, and M. A. Schweisfurth, "Closed-Loop Multi-Amplitude control for robust and dexterous performance of myoelectric prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, pp. 498-507, March 2020.
- [6] Q. Nie, J. Wang, and X. Wang, "View-invariant human action recognition based on a 3d bio-constrained skeleton model," *IEEE Trans. Image Process.*, vol. 28, pp. 3959-3972, July 2019.
- [7] M. Nieto-Hidalgo, F. J. Ferrández-Pastor, and R. J. Valdivieso-Sarabia, "A vision based proposal for classification of normal and abnormal gait using RGB camera," *J. Biomed. Inform.*, vol. 63, pp. 82-89, January 2016.
- [8] W. Min, L. Yao, and Z. Lin, "Support vector machine approach to fall recognition based on simplified expression of human skeleton action and fast detection of start key frame using torso angle," *IET Comput. Vis.*, vol. 12, pp. 1133-1140, December 2018.
- [9] A. A. Liu, W. Z. Nie, and Y. T. Su, "Coupled hidden conditional random fields for RGB-D human action recognition," *Signal Process.*, vol. 112, pp. 74-82, July 2015.
- [10] Khatun M A, Yousuf M A, Ahmed S, et al. "Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor," *IEEE J. Transl. Eng. Health Med.-JTEHM*, vol. 10, pp. 1-16, May 2022.
- [11] L. J. Zhang. "Research on human body movement posture based on inertial sensor," *Int. J. Bioautomation*, vol. 22, pp. 179-186, July 2018.
- [12] J. J. Marron, M. A. Labrador, and A. Menendez-Valle, "Multi sensor system for pedestrian tracking and activity recognition in indoor environments," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 23, pp. 3-23, March 2016.
- [13] M. Guo, Z. L. Wang. "Segmentation and recognition of human motion sequences using wearable inertial sensors," *Multimed. Tools Appl.*, vol. 77, pp. 21201-21220, August 2018.
- [14] J. X. Qi, G. Z. Jiang, and G. F. Li, "Surface EMG hand gesture recognition system based on PCA and GRNN," *Neural Comput. Appl.*, vol. 32, pp. 6343-6351, May 2020.
- [15] Zhang P, Zhang J, Elsabbagh A. "Lower Limb Motion Intention Recognition Based on sEMG Fusion Features," *IEEE Sens. J.*, vol. 22, pp. 7005-7014, January 2022.
- [16] Q. Ye, P. Huang, and Z. Zhang, "Multiview learning with robust double-sided twin SVM," *IEEE T. Cybern.*, vol. 52, pp. 12745-12758, September 2021.
- [17] V. A. Binson, M. Subramoniam, and Y. Sunny, "Prediction of pulmonary diseases with electronic nose using SVM and XGBoost," *IEEE Sens. J.*, vol. 21, pp. 20886-20895, September 2021.
- [18] M. Awais, X. Long, and B. Yin, "A hybrid DCNN-SVM model for classifying neonatal sleep and wake states based on facial expressions in video," *IEEE J. Biomed. Health Inform.*, vol. 25, pp. 1441-1449, May 2021.
- [19] A. H. Gandomi, X. S. Yang, and A. H. Alavi, "Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems," *Eng. Comput.*, vol. 29, pp. 1-19, January 2013.
- [20] H. Osman, M. F. Baki. "A cuckoo search algorithm to solve transfer line balancing problems with different cutting conditions," *IEEE Trans. Eng. Manag.*, vol. 65, pp. 505-518, July 2018.

# Bidirectional Recurrent Neural Network based on Multi-Kernel Learning Support Vector Machine for Transformer Fault Diagnosis

Xun Zhao<sup>1</sup>, Shuai Chen<sup>2\*</sup>, Ke Gao<sup>3</sup>, Lin Luo<sup>4</sup>

School of Information and Control Engineering, Liaoning Petrochemical University, Fushun, China<sup>1,2,4</sup>  
Offshore Oil Engineering CO.LTD, Tianjin, China<sup>3</sup>

**Abstract**—Traditional neural network has many weaknesses, such as a lack of mining transformer timing relation, poor generalization of classification, and low classification accuracy of heterogeneous data. Aiming at questions raised, this paper proposes a bidirectional recurrent neural network model based on a multi-kernel learning support vector machine. Through a bidirectional recurrent neural network for feature extraction, the features of the before and after time fusion and obvious data are outputted. The multi-kernel learning support vector machine method was carried out on the characteristics of data classification. The study of multi-kernel support vector machines in the weighted average of the way nuclear fusion improves the accuracy of characteristic data classification. Numerical simulation analysis of the temporal channel length for sequential network diagnostic performance, the effects of multi-kernel learning on the generalization ability of support vector machine, the influence on heterogeneous data processing capabilities, and transformer fault data classification experiment verifies the correctness and effectiveness of the bidirectional recurrent neural network based on multi-kernel learning support vector machine model. The experiment result shows that the diagnosis performance of bidirectional recurrent networks based on a multi-kernel learning support vector machine is better, and the prediction accuracy of the model is improved by more than 1.78% compared with several commonly used neural networks.

**Keywords**—Multi-kernel learning; support vector machine; bidirectional recurrent neural network; fault diagnosis

## I. INTRODUCTION

A transformer, one of the key hub equipment in the power grid, act as an important link in energy conversion, distribution, or transmission. Transformer failure will cause huge financial loss and endanger public security. Therefore, timely and accurate diagnosis of transformer failure has important significance to make sure that the power system in safe state [1]. Characteristic gas method is often used for manual diagnosis of traditional transformer faults, but the differences in experience for fault identification often lead to errors. For transformer insulation maintenance, manual judgment requires not only a power failure of tested transformers but also regular maintenance of replaced equipment or parts. And a large part of equipment does not exceed its service life, which often causes a waste of resources and reduces the economy of the power system.

When the LSTM network conducts data analysis on time series feature quantity, the model is too simple and can only

consider a single time series. The fault data identification is not ideal and the accuracy of long-time series will decline. Besides, the ability to generalize is not high. In this paper, a Bidirectional Long Short-Term Memory (Bi-LSTM) network is proposed and Multi-Kernel Learning Support Vector Machines (MKL-SVM) combined optimization algorithm. In this method, a reverse sequential LSTM network is added to establish a Bi-LSTM network model, which can consider the feature quantity of the future time. Then, the kernel functions used by SVM are aggregated together by the weighted average method to establish the MKL-SVM model, and MKL-SVM is used to replace the Softmax function to achieve fault classification. The feature of SVM is applied to increase the classification generalization ability, and multi-kernel learning has more excellent characteristics for heterogeneous data classification. A failure in the transformer oil dissolved gas H<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>4</sub>, C<sub>2</sub>H<sub>2</sub>, the characteristics of the gas is chosen to verify the effectiveness of multi-kernel learning support vector machine and bidirectional recurrent neural network for transformer fault diagnosis. Compared with several existing prediction models, the model can make a more accurate judgment on a few heterogeneous data, with stronger generalization ability and higher prediction accuracy.

The rest of this paper is consisted of as: Section II presents the related works. The proposed method is described in Section III and Section IV. Experimental results and performance analysis are discussed in Section V. Finally, this paper concludes in Section VI.

## II. RELATED WORK

Machine learning is used increasingly frequently and effectively to process and analyze data as a result of technological advancements. Zhang Hang et al. proposed to use support vector machine to improve the accuracy of motor fault diagnosis, but the model has many limitations, including poor generalization ability, low classification efficiency and high interference accuracy [2]. Zhang Xin, Wang Heng et al. come up with a model of a neural network optimized by a sparrow search algorithm, which in view of transformer fault diagnosis to increase the correctness more effectively. However, the number of probabilistic neural network neurons was affected by training samples, and the time it takes to train the model would be increased to get a better model [3]. For increase the stability of SVM, Qingchuan Fan come up with a whale optimization algorithm that introduced the

characteristics of a genetic algorithm to improve SVM but sacrificed part of the classification accuracy of SVM [4]. In view of the contradiction between the stability and accuracy of SVM, Yuhan Wu et al. proposed an SVM optimization algorithm that adopts adaptive probability formula to balance the squirrel search algorithm. The iteration cycle and accuracy of the model are increased [5]. Bing Zeng et al. proposed an optimized gray wolf algorithm Let-Squares SVM combining particle swarm optimization and differential evolution. The model got rid of the weakness which it was caught in local optimum frequently and improved the correct rate of the model, but the generalization ability of the model was greatly affected [6].

The emergence of deep learning brings new changes to machine learning. Miao Jianjie et al. proposed to use improved particle swarm optimization algorithm to optimize fuzzy neural network, automatically adjust parameters and accelerate convergence [7]. Gao Xincheng et al. proposed using improved genetic algorithm to optimize convolutional neural networks, which shortened the time to obtain the optimal weight and improved the convergence and accuracy of neural networks [8]. With the emergence of more complex neural networks, more deep-learning methods have been put into use for fault diagnosis. Taha Ibrahim B. M. et al. come up with an improved convolutional neural network suitable for noise environment, which perfected the accuracy of the Convolutional Neural Network (CNN) model. But the CNN model does not have a good classification effect on time series data and cannot effectively extract features from time series data [9]. In order to enable neural networks to better consider timing features, Fan Xiaodong et al. come up with apply Long Short-Term Memory (LSTM) network for transformer fault diagnosis, which has a better effect than CNN. When LSTM is applied to fault diagnosis, it has trouble to ensure the stability of the model and the accuracy of long-time series will decline as the number of training increases [10]. To improve the stability of the model, He Yigang and Wu Xiaoxin put forward a kind of complicated correlation characteristic of bidirectional Recurrent Neural Network (RNN). The stability of the model is higher and can be related to more features. However, when we use models, the ability to generalize is not high with only the specific training being set to get a better training result [11]. Omar Alharbi combine CNN and Bi-LSTM for generating the final features representation to be passed to a linear SVM classifier [12]. This method used for classification of Arabic reviews show that the method achieved superior performance than the two baseline algorithms of CNN and SVM in all datasets. The combination of multiple machine learning methods has a better development prospect.

### III. FAULT DIAGNOSIS PRINCIPLE OF TRANSFORMER WITH BIDIRECTIONAL RECURRENT NEURAL NETWORK

Dissolved Gas Analysis (DGA) in oil is often applied to transformer fault diagnosis, which has the merit of easy to operate and strong anti-interference ability. In order to ensure reliable data collection and effective analysis results, the DGA method is adopted in this paper. DGA is a series of classical time series data, which collects dissolved gas data in transformer oil in a fixed period without interruption in time sequence. Therefore, the bidirectional recurrent neural network can be used as the feature extraction part of the model to feature extraction of time sequence data in transformer oil. Make sure that the timing sequence feature extraction ability of RNN and the dependability of the real operation, in the passage adopts an LSTM network with better timing sequence extraction ability and stable operation.

#### A. Dissolved Gas Analysis Method in the Transformer Oil

Transformer faults are split into electrical faults, thermal faults, mechanical faults. Yet as the frequency of mechanical faults is the lowest with the rest of the several faults, the main analysis object of thermal faults and electrical faults.

When transformer is in the process of fault operation, the insulation oil of the transformer will be oxidized and cracked due to the action of discharge and heat. The main composition of the insulation oil hydrocarbon will produce hydrogen and low molecular alkenes, alkenes, alkynes, and other gases. With the severity of the fault, the rising rate of each gas is different, and the transformer fault type can be roughly judged according to the different types of gas and concentration. Table I lists the characteristic gases of different faults.

TABLE I. THE CHARACTERISTIC GASES OF DIFFERENT FAULTS

Fault Types	Main Characteristic Gases	Accompany Characteristic Gases
Oil overheating	Methane CH <sub>4</sub> , Acetylene C <sub>2</sub> H <sub>2</sub>	Hydrogen H <sub>2</sub> , Ethane C <sub>2</sub> H <sub>6</sub>
Oil and paper overheating	Methane CH <sub>4</sub> , Ethylene C <sub>2</sub> H <sub>4</sub> , Carbon monoxide CO	Hydrogen H <sub>2</sub> , Ethane C <sub>2</sub> H <sub>6</sub> , Carbon dioxide CO <sub>2</sub>
Partial discharge in oil paper insulation	Hydrogen H <sub>2</sub> , Methane CH <sub>4</sub> , Carbon monoxide CO	Ethylene C <sub>2</sub> H <sub>4</sub> 、Ethane C <sub>2</sub> H <sub>6</sub> 、Acetylene C <sub>2</sub> H <sub>2</sub>
Spark discharge in oil	Hydrogen H <sub>2</sub> , Acetylene C <sub>2</sub> H <sub>2</sub>	
Arc in oil	Hydrogen H <sub>2</sub> , Acetylene C <sub>2</sub> H <sub>2</sub> , Ethylene C <sub>2</sub> H <sub>4</sub>	Methane CH <sub>4</sub> , Ethane C <sub>2</sub> H <sub>6</sub>

According to Table I, the characteristics of gas under different fault types, the transformer insulating oil by electrolysis produces several kinds of characteristic gas for parameters. And the characteristics of gas concentration and velocity were analyzed in different operation conditions while the operating state of the transformer was assessed to decision the transformer fault types. The transformer oil is a dissolved gas analysis method. This analysis method has the advantage of supporting live online detection to spare it from being impacted by the signal of electric and magnetic fields and has a simple operation mode. DGA applied in transformer state monitoring and fault diagnosis [13-14].

H<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>4</sub>, C<sub>2</sub>H<sub>2</sub>, C<sub>2</sub>H<sub>6</sub>, and the other five gases in the gas concentration detection results are more accurate and can determine the transformer fault type. In this experiment, H<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>4</sub>, C<sub>2</sub>H<sub>2</sub>, C<sub>2</sub>H<sub>6</sub> were chosen as the characteristic gases of experiment by using transformer fault data. Fig. 1 shows the concentration curves of the five characteristic gases.

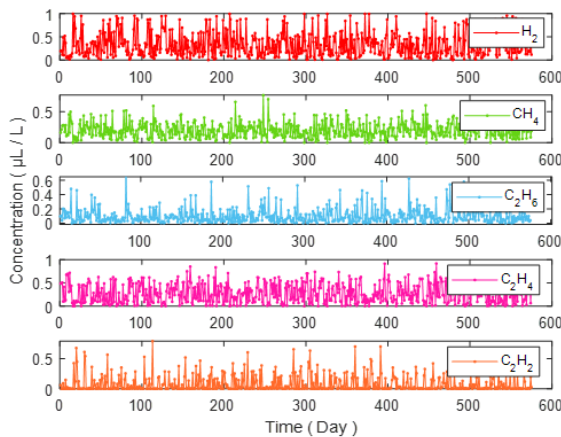


Fig. 1. The concentration curves of the five characteristic gases

### B. Long Short-Term Memory Network

The perfect forecast of dissolved gas in oil can more effectively understand the operation state of the transformer and make a timely judgment of the fault. The Neural structure of a Recurrent Neural Network (RNN) has a self-feedback function, which can retain both current and previous information at the same time and can be used to calculate the current output [15]. However, when the RNN model is used to analyze the long time series data, there are defects of gradient explosion or gradient disappearance present to delayed backpropagation during the training process. So the RNN network is not good at analyzing long-series data [16]. LSTM network is a kind of RNN model that can store time sequence information for a long time by adding a gating unit on the basis of a general recurrent neural network. As a way of deep learning, the recurrent neural network can extract data features from time series more efficiently and accurately. The quality of features determines the accuracy of classification, and accurate classification of fault data can make a more timely response to transformer faults. For increasing the precision rate of neural network feature extraction for time series data, Take LSTM model as the prototype, the reverse time-series memory

network was introduced to increase the dependence of characteristic gases on time series information and increase the precision rate gas prediction by the model.

The basic unit structure of the LSTM is shown in Fig. 2 is the input of the current moment while being the value of the previous moment in memory unit. The activation function  $\sigma$  usually used the sigmoid function. Compared with RNN, LSTM is characterized by a gating mechanism, which includes three parts, namely the forget gate, input gate, and output gate [17]. The responsibility of forget gate is selectively forgetting the historical information stored in the memory unit. The input gate retains the current external information and integrates it with the historical information. And the output gate decides whether to output. The output of the LSTM is determined jointly by the gating unit and the input.

$$f_t = \sigma(W_f[X_t; Y_{t-1}] + b_f) \quad (1)$$

$$i_t = \sigma(W_i[X_t; Y_{t-1}] + b_i) \quad (2)$$

$$o_t = \sigma(W_o[X_t; Y_{t-1}] + b_o) \quad (3)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot g(X_t) \quad (4)$$

$$Y_t = o_t \cdot h(c_t) \quad (5)$$

In the above equation,  $W_f$  and  $b_f$ ,  $W_i$  and  $b_i$ ,  $W_o$  and  $b_o$  are weight and bias for the forget gate, input gate, output gate respectively.  $g(\cdot)$  and  $h(\cdot)$  are mappings of  $X_t \rightarrow c_t$  and  $c_t \rightarrow Y_t$  respectively represents matrix multiplication.  $c_t$  contains both the input state and the history state, which improves the stability of the neural network and solves the problem of gradient disappearance that often occurs in the RNN.

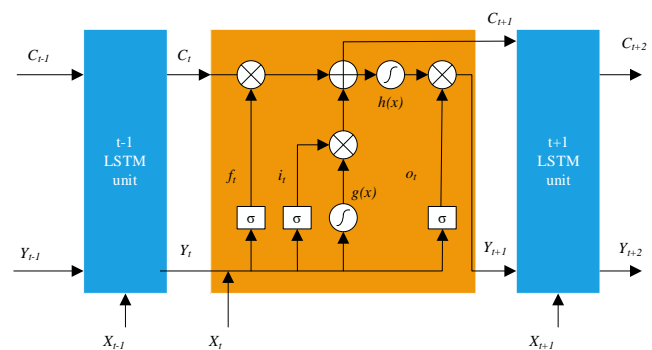


Fig. 2. The time sequence diagram of LSTM

The time sequence diagram of LSTM is shown in Fig. 2. In each epoch of LSTM, the output will affect the gating unit so that the information will not be lost. Because of the existence of the forget gate, LSTM can decide when to discard some unimportant time information, to maintain the updating speed and certain accuracy of the network in the long time series.

The dissolved gas data in transformer oil is typical time series data. The data at a certain moment is closely related to the data at the previous moment, and the general neural network cannot find this connection, thus limiting the accuracy of the model prediction. LSTM can store historical data at a certain time scale due to the existence of memory units. This structure gives LSTM great advantages in processing time series feature data.

#### IV. BIDIRECTIONAL RECURRENT NEURAL NETWORK BASED ON MULTI-KERNEL LEARNING SUPPORT VECTOR MACHINE DIAGNOSIS MODEL

Generally, the recurrent neural network uses Softmax function to classify, but classification effect depends on the data analysis of recurrent neural network and the ability of generalization is bad. For solving this problem, we use SVM to replace the Softmax function in a recurrent neural network. Dissolved gas in transformer oil analysis by recurrent neural network, there are often some heterogeneous data located at the edge of classification space. For single-kernel support vector machines, it is difficult to process heterogeneous data, and they often ignore the heterogeneous data. For making more effective use of the processed heterogeneous data, MKL-SVM is produced by combining multiple kernel functions. The bidirectional recurrent neural network model based on a multi-kernel learning support vector machine is composed of two parts. Feature extraction based on a bidirectional recurrent neural network is an important part. The classifier of the second part uses MKL-SVM.

##### A. Data Preprocessing

According to the selection of H<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>4</sub>, C<sub>2</sub>H<sub>2</sub>, and C<sub>2</sub>H<sub>6</sub>, the five characteristics gases, and the ratio of the total volume input of the model, the data normalization processing is carried out according to the (6).

$$x_m^{(i)} = \frac{x^i}{x^1 + x^2 + x^3 + x^4 + x^5}, i = 1, 2, 3, 4, 5 \quad (6)$$

In the above equation,  $x_m^{(i)}$  is normalized.  $x^i (i = 1, 2, 3, 4, 5)$  is the volumes of H<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>4</sub>, C<sub>2</sub>H<sub>2</sub>, and C<sub>2</sub>H<sub>6</sub> before normalization respectively.

##### B. Training Set Selection

For the volume fraction set of dissolved gas in oil, the K-fold cross-validation method is adopted to partition the data set into training and test two parts. The cross-validation times are 5. The training set is responsible for training model, and the accuracy is evaluated through the test set, which can avoid the phenomenon of over-fitting. For large data sets, a relatively small training set can meet the requirements of model training. For small data set, a relatively large training set is needed to train model. In this experiment, the size of data set is small, so training set and test set account for 80% and 20% each.

Based on the study of various representative characteristic gases of transformer faults in Table I, combined with the faults frequently encountered in actual production, a variety of faults are summarized according to the characteristics of faults to deal with faults faster in actual operation and optimized

classification ability. And the operation states of transformers are classified into six categories: (1) Normal, (2) partial discharge, (3) low energy discharge, (4) high energy discharge, (5) medium-low temperature overheating, and (6) high temperature overheating.

##### C. Bidirectional Circulation Structure

Generally, the data analysis of recurrent neural networks depends on time series. Recurrent neural network need the data of the previous moment to predict the change of future data. When data has a temporal relationship or the data is more time-dependent, such network structure will often produce large errors. In addition, the training of this kind of neural network requires large-scale training samples. If there are few elements in the training sample, it may be difficult to obtain an ideal model.

To solve these problems, a bidirectional recurrent neural network is adopted in this paper. On the basis of original recurrent neural network model, a reverse sequential recurrent neural network is added. Two recurrent neural networks share input and output. Fig. 3 shows the structure of the bidirectional recurrent neural network.

As Fig. 3 show, the recurrent neural network has two hidden layers. The data is simultaneously input into the forward and reverse timing sequence. W<sub>2</sub> is the forward timing sequence, and W<sub>5</sub> is the reverse timing sequence. W<sub>1</sub> enters the forward timing sequence, and W<sub>3</sub> enters the reverse timing sequence. W<sub>4</sub> and W<sub>6</sub> jointly constitute the output layer. Different from the general neural network, the forward cycle layer and reverse cycle layer in the recurrent neural network are not connected, which can effectively prevent the occurrence of the self-cycle phenomenon.

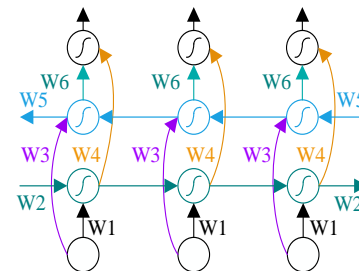


Fig. 3. The structure of the bidirectional recurrent neural network

##### D. Support Vector Machine

The recurrent neural network uses the Softmax function in fault classification process, but this function cannot perform well in a wide range of faults. When the Softmax function is multi-classified, its effect is similar to the Sigmoid function used for Logistic regression. Only when various types are mutually exclusive, can the Softmax function have good classification ability [18]. SVM has a good classification effect and generalization ability in the classification task and it is a common fault diagnosis method to combine data analysis and classification with a recurrent neural network. The LSTM-SVM model obtained by combining the LSTM network with the SVM classifier can simultaneously give full play to LSTM's ability to process long sequence information and SVM's ability to classify low-dimensional feature data, to

reduce modeling difficulty and computational complexity and improve the speed and accuracy of performance degradation prediction [19].

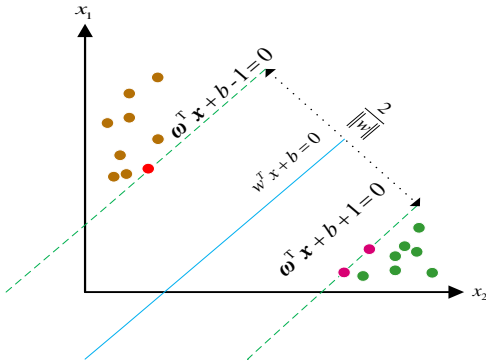


Fig. 4. The diagram of the hyperplane dividing two types of samples

The central theme of the SVM is classification learning. For a given training sample set, a classification hyperplane is found in the sample space as the decision boundary to separate samples of different categories [20]. Fig. 4 shows the diagram of the hyperplane dividing two types of samples.

The hyperplane described by the following linear equation:

$$0 = \omega^T \mathbf{x} + b \quad (7)$$

Sample points in the sample space satisfy the following equation:

$$y = \omega^T \mathbf{x} + b \quad (8)$$

In the above equation,  $b$  is the bias of the value,  $\mathbf{x}$  is the data set of the input,  $\omega$  is the classification hyperplane of the partition class.

The distance that sample arrive the hyperplane in the sample space expressed as follows:

$$r = \frac{|\omega^T \mathbf{x} + b|}{\|\omega\|} \quad (9)$$

The nearest sample that arrive the hyperplane in the sample space is called 'support vector', and the sum of distance between the support vectors of two classes and the hyperplane is as follows:

$$\gamma = \frac{2}{\|\omega\|} \quad (10)$$

The basic form of a support vector machine can be obtained by maximizing the geometric interval between the data set and hyperplane.

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 \quad (11)$$

$$s.t. y_i(\omega^T \mathbf{x}_i + b) \geq 1, i = 1, 2, \dots, m.$$

From (11), the solution of SVM is a convex quadratic programming problem, duality problem got by using the

Lagrangian multiplier method. The Lagrangian function of the dual problem can be expressed by the following formula.

$$L(\omega, b, \alpha) = \frac{1}{2} \|\omega\|^2 + \sum_{i=1}^m \alpha_i (1 - y_i(\omega^T \mathbf{x}_i + b)) \quad (12)$$

Here,  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$ . The partial derivatives of the Lagrangian function  $L$  are taken with respect to  $\omega$  and  $b$ , and they are set to zero respectively to get the optimal solution of the function.

$$\omega = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i \quad (13)$$

$$0 = \sum_{i=1}^m \alpha_i y_i \quad (14)$$

According to (11), the dual problem is obtained by substituting (13) and (14) into (12):

$$\max_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (15)$$

$$s.t. \sum_{i=1}^m \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \dots, m$$

According to (8), (13), and (14), the model is obtained.

$$f(\mathbf{x}) = \omega^T \mathbf{x} + b = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \quad (16)$$

It can be seen from (11) that there are inequality constraints on  $\alpha_i$  and training samples  $(\mathbf{x}_i, y_i)$  in support vector machines.

$$\begin{cases} \alpha_i \geq 0 \\ y_i f(\mathbf{x}_i) - 1 \geq 0 \\ \alpha_i (y_i f(\mathbf{x}_i) - 1) = 0 \end{cases} \quad (17)$$

The above constraints stipulate that the training samples always have  $\alpha_i = 0$  or  $y_i f(\mathbf{x}_i) = 1$ .

Thinking about nonlinear case, the hyperplane model in the characteristic space is calculated below:

$$f(\mathbf{x}) = \omega^T \varphi(\mathbf{x}) + b \quad (18)$$

From (18),  $\varphi(\mathbf{x})$  is the expression generated after the two-dimensional sample  $\mathbf{x}$  to a higher dimensional feature space. According to (12), (13), and (14), a new dual problem is as follows:

$$\max_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j) \quad (19)$$

$$s.t. \sum_{i=1}^m \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \dots, m$$

It is necessary to find a  $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j)$  function, namely the kernel function of SVM, which generally uses

linear kernel, Gaussian function kernel, polynomial kernel, and sigmoid kernel.

### E. Multi-Kernel Learning

Generally, single-kernel structures are used in support vector machines, which are classified based on a single feature space. The selection of kernel functions needs to be judged according to actual needs and then selected according to experiences. And different parameters are set. Such kernel function design is not convenient, and SVM is difficult to train the specific data in the training sample, resulting in lower accuracy of the classifier.

For solving above flaws of SVM, this paper adopts the combination of multiple kernel functions to establish the MKL-SVM model and improves the adaptability of support vector machine to complex data through multi-kernel learning. MKL integrates multiple sub-kernels into a unified optimization framework to seek the best combination. Using the multi-kernel model optimize the performance of the learning model and obtain explicable decision functions [21]. With the adoption of the multi-kernel learning method, the classification effect of a single support vector machine on a few specific data is changed, and the classification accuracy of fault data is improved.

The kernel function in multi-kernel learning is composed of several basic kernel functions.

$$K'(x_i, x_j) = \sum_{i=1}^m \lambda_i K_i(x_i, x_j), \lambda_i \geq 0; \sum_{i=1}^m \lambda_i = 1 \quad (20)$$

From (20),  $m$  is the number of kernels, and  $\lambda_i$  is the weight of each kernel. Before calculating the weight of each kernel function, we need to obtain the kernel matrix of each kernel function.

$$K_i = \begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_m) \\ \vdots & \ddots & \vdots \\ k(x_m, x_1) & \dots & k(x_m, x_m) \end{bmatrix}, i = 1, 2, \dots, m \quad (21)$$

After the kernel matrix is obtained, the eigenmatrix  $M$  is calculated according to the eigenvectors that get by data set. Getting the weight of each kernel function, the trace of each matrix  $\text{tr}(K)$  and  $\text{tr}(M)$  is used to characterize the characteristics of each matrix. And then the Euclidean distance of each  $\text{tr}(K)$  and  $\text{tr}(M)$  is calculated as follows.

$$|X_i| = \sqrt{\text{tr}(K_i) + \text{tr}(M)}, i = 1, 2, \dots, m \quad (22)$$

Then the kernel matrix is substituted into the feature matrix to obtain the importance of each kernel matrix to the feature.

$$H_i = \frac{K_i M}{|X_i|} \quad (23)$$

Finally, the weights of each kernel function are obtained.

$$\lambda_i = \frac{H_i}{\sum_{i=1}^m H_i} \quad (24)$$

### F. Calculation Steps of the Fault Diagnosis Method

The flow chart of Bidirectional Long Short-Term Memory network based on Multi-Kernel Learning Support Vector Machine (Bi-LSTM-MKL-SVM) is shown in Fig. 5.

Sum the output values at the current time to get  $Y_t$ .

$$Y_t = Y_t^1 + Y_t^2 \quad (25)$$

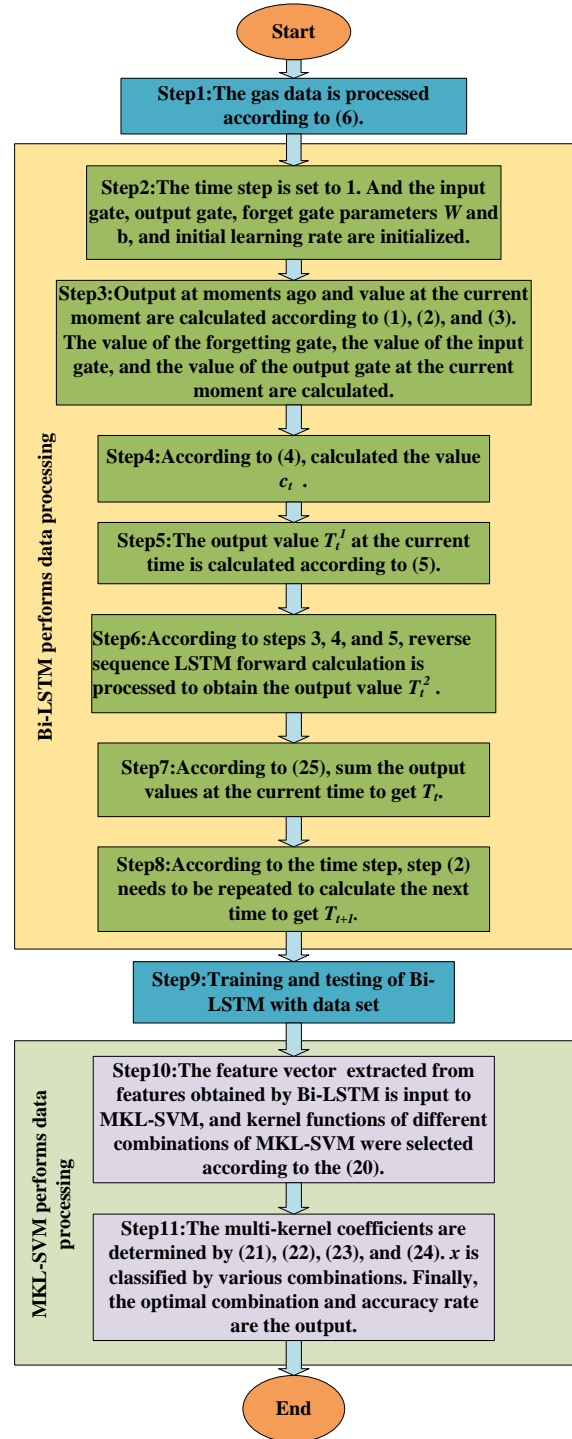


Fig. 5. The flow chart of Bi-LSTM-MKL-SVM

## V. EXPERIMENTAL ANALYSIS

### A. Selection of Training Times

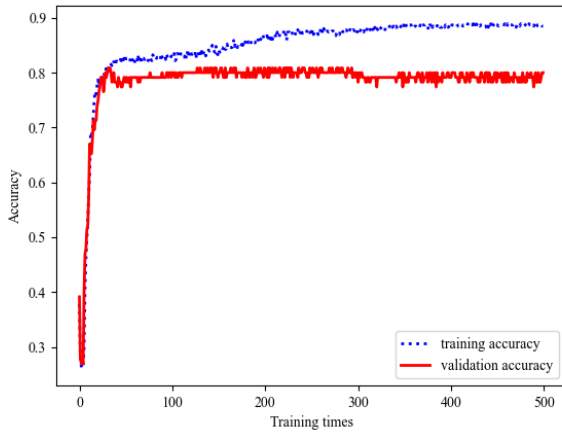


Fig. 6. The accuracy of the training set and test set of Bi-LSTM

The accuracy of the training set and test set of Bi-LSTM is shown in Fig. 6. After 300 training times, the accuracy of the test set decreases obviously, and the model training has an overfitting phenomenon. Thus the number of sessions is set to 300 for the best results.

### B. Feature Extraction Performance Analysis of Time Series Data

In Table II, the model related parameters for using method in this paper are presented. It mainly includes the model mentioned in the paper and the parameters used in the model.

TABLE II. DESCRIPTION OF MODEL RELATED PARAMETERS

Abbreviation	Description
Bi-LSTM	It represents an LSTM with an additional reverse sequence
Bi-LSTM-SVM	Bi-LSTM is connected with SVM, and SVM is used to replace the classification function in Bi-LSTM
MKL-SVM	SVM is used by stacking multiple kernel functions according to a certain weight
Bi-LSTM- MKL-SVM	Bi-LSTM is connected with MKL-SVM, and MKL-SVM is used to replace the classification function in Bi-LSTM
Time step	The step size of each computation in LSTM
forget gate	It is responsible for controlling the persistence of memory cell
input gate	It is responsible for controlling the input of the immediate state into memory cell
output gate	It is responsible for controlling whether the current value is printed
Memory cell	It is responsible for storing long-term state

In order to validate the bidirectional recurrent neural network to improve the effectiveness of time-series analysis, training set input Bi-LSTM respectively with LSTM, RNN, and CNN model and then validated the accuracy of test set. The results of four kinds of models of time-series data feature extraction accuracy are shown in Table III.

TABLE III. THE RESULTS OF FOUR KINDS OF MODELS OF TIME-SERIES DATA FEATURE EXTRACTION ACCURACY

Table No	Models	Accuracy (%)
1	CNN	80.18
2	RNN	81.34
3	LSTM	82.83
4	Bi-LSTM	84.08

To ensure no interference to the sequential feature extraction ability of the models, the same Softmax function was used as the classifier for the four models. According to the data in Table III, RNN has higher accuracy than CNN in time sequence data because of its memorability. Compared with RNN, the LSTM network has higher accuracy in data analysis due to the gated unit, while RNN has the problem of gradient explosion due to the rugged search path. So pruning operation is required to decrease learning rate and increase training time. Among above recurrent neural networks, Bi-LSTM has the highest accuracy, which is more than 84%. In the process of model training, Bi-LSTM model not only speeds up the convergence rate of the model but also deepens the time sequence at the same time for the correlation of each parameter. It also corrects the training model, so that the model has better accuracy in prediction. The volatility of its accuracy is less than other recurrent neural networks and CNN. The experiment proves that the Bi-LSTM network can better complete the feature extraction task in time series data by adding a reverse time series recurrent neural network. MKL-SVM has a better classification effect on heterogeneous data. By comparing the confusion matrix between CNN and Bi-LSTM, we verify whether the heterogeneous data generated by Bi-LSTM is more suitable for MKL-SVM on the basis of Bi-LSTM having a better feature extraction ability on time series data. SVM, as a classical classifier, is compared with Bi-LSTM to verify the influence of feature extraction of time series data on data sample classification. The confusion matrix of CNN, Bi-LSTM, and SVM is shown in Fig. 7.



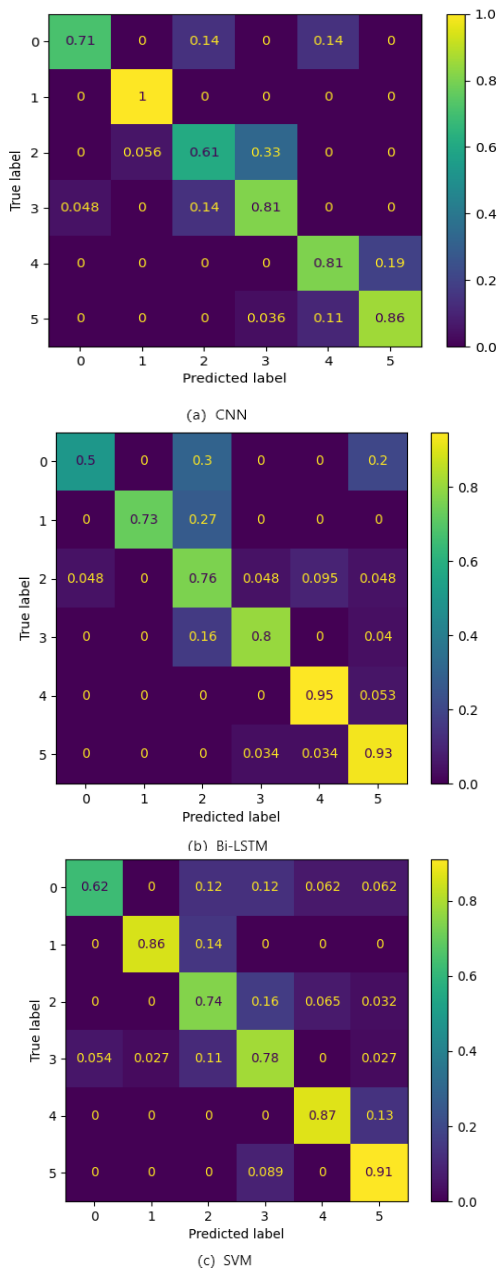


Fig. 7. The confusion matrix of CNN, Bi-LSTM, SVM

Both CNN and Bi-LSTM use the Softmax function as model classifier. CNN model only has a good classification effect on partial discharge fault type, but a poor classification effect on other fault types. The Bi-LSTM model has a good classification effect on multiple fault types. According to Fig. 7, the distribution range of Bi-LSTM classification results is wider, indicating that Bi-LSTM will generate more heterogeneous data in feature extraction of data and affect the classification effect of the classifier. SVM and Bi-LSTM have

similar fault diagnosis effects, but SVM has more misjudgment of fault types and the classification effect is not stable, indicating that the data extracted by feature can be more easily classified. The confusion matrix shows the classification performance of Bi-LSTM is significantly better than CNN and SVM. After feature extraction, the Bi-LSTM time series data generates more heterogeneous data and has a better combination with MKL-SVM.

### C. Diagnosis Result Analysis of Bidirectional Recurrent Neural Network Model based on a Multi-Kernel Support Vector Machine

In this section, MKL-SVM model will be established through kernel fusion to form a multi-kernel. In multi-kernel learning, multiple kernel functions can be used for fusion, or the same kernel function with different parameters can be used for fusion to ensure the diversity of kernel fusion. Fig. 8 shows the workflow of MKL-SVM.

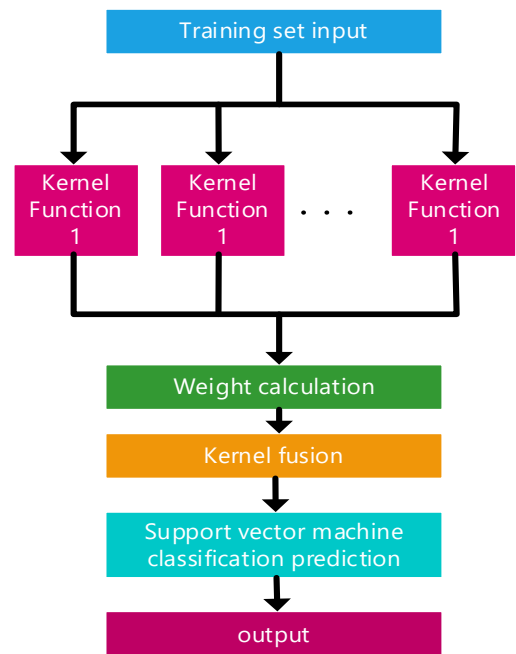


Fig. 8. The workflow of MKL-SVM

To prove the effectiveness of Bi-LSTM-MKL-SVM network in improving the exactitude of time series analysis and prediction, SVM and Bidirectional Long Short-Term Memory network based on Support Vector Machines (Bi-LSTM-SVM), LSTM and Bi-LSTM and Bi-LSTM-MKL-SVM are used to compare and analyze the predicted results of test sets.

The micro-average PR curve of the five models is shown in Fig. 9. On the whole, under the influence of heterogeneous data, the classification performance of Bi-LSTM-SVM is not different from Bi-LSTM and LSTM. Bi-LSTM-MKL-SVM has better classification performance, while the SVM has the worst classification effect.

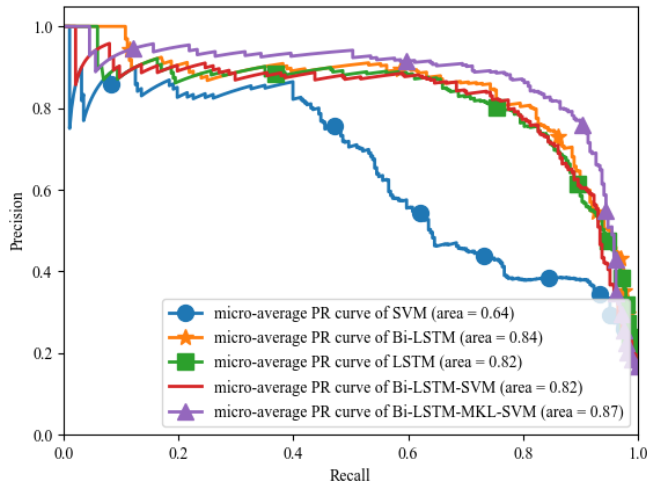


Fig. 9. The micro-average PR curve of the five models

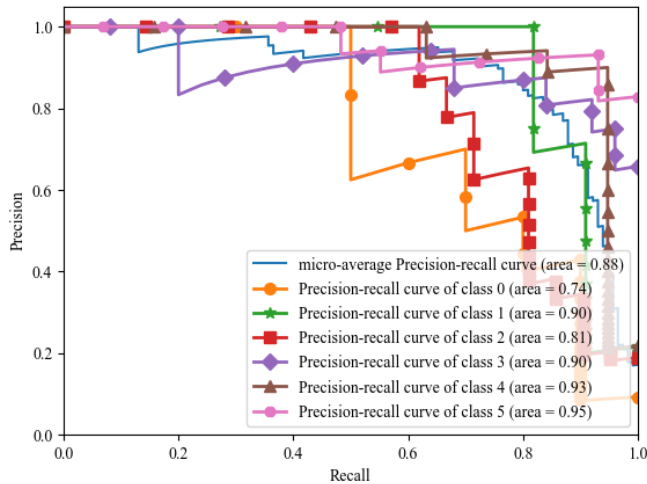
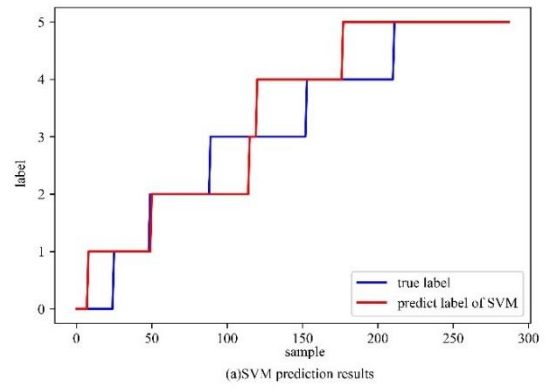
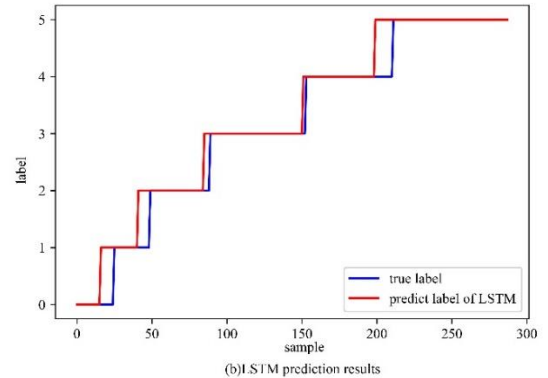


Fig. 10. The PR curve of the Bi-LSTM-MKL-SVM model

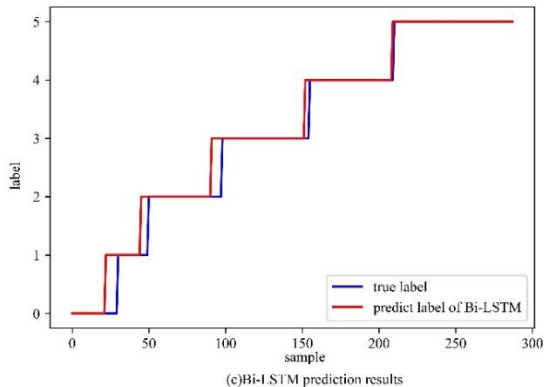
The PR curve of the Bi-LSTM-MKL-SVM model is shown in Fig. 10. Bi-LSTM-MKL-SVM has good classification performance for each fault category under good overall classification performance. Compared with the SVM model, the combined model that extract time series information by the recurrent neural network can effectively improve the classification efficiency of the classifier. For the LSTM model, SVM classification can significantly optimize neural network performance. Feature extraction of time series data by Bi-LSTM will generate a lot of heterogeneous data. Using general SVM can not only improve the performance of the classifier but also cause interference with the classification ability of SVM. MKL-SVM is more effective for the classification of heterogeneous data. Bi-LSTM-MKL-SVM is used to solve the influence of heterogeneous data on model judgment better, improve the model's utilization ability of time-series data, and enhance the overall generalization ability of model, as well as improves its stability.



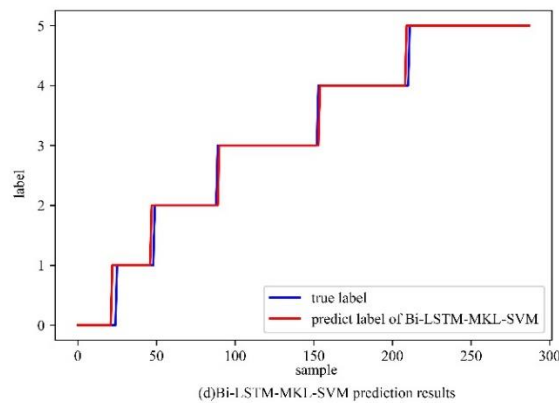
(a)SVM prediction results



(b)LSTM prediction results



(c)Bi-LSTM prediction results



(d)Bi-LSTM-MKL-SVM prediction results

Fig. 11. The comparison between the true label and predicted label

Fig. 11 shows the comparison between the true label and the predicted label. The data fitting and classification effect are shown in Fig. 11; the SVM model has a poor fitting effect on test samples. This shows that simple classification algorithm is difficult to complete accurate classification of complex data. The Bi-LSTM model has better fitting accuracy than the LSTM model. This indicates that adding a reverse timing network can effectively improve the accuracy of LSTM. Compared with the other three models, the Bi-LSTM-MKL-SVM model has the best classification and fitting effect on test samples, indicating that the Bi-LSTM-MKL-SVM model has a stronger ability to extract time series features. It proves that these optimizations are effective. SVM can significantly improve the generalization of neural networks. Multi-kernel learning can effectively optimize the classification effect of SVM on complex data.

TABLE IV. THE ACCURACY OF THE FIVE MODELS

Table No.	Models	Accuracy (%)
1	SVM	81.50
2	LSTM	82.83
3	Bi-LSTM-SVM	83.58
4	Bi-LSTM	84.08
5	Bi-LSTM-MKL-SVM	85.86

The accuracy of the five models is shown in Table IV. The Bi-LSTM network model is used to process time series data, which improves the classification efficiency of MKL-SVM, improves the accuracy and stability of the model as a whole, and effectively solves the shortcomings of neural networks and SVM used alone. It can be seen by comparing with the model prediction accuracy, MKL-SVM does better in heterogeneous data for more accurate judgment, and it can save time for SVM to adjust the parameter and reduce the possibility of error when choosing parameters based on experience. More basic kernel function combinations optimize the effect of MKL-SVM and its applicability.

## VI. CONCLUSION

In allusion to transformer fault diagnosis, the deep learning method has problems with the stability and the effective use of data samples. This paper comes up with a kind of model called Bi-LSTM-MKL-SVM and takes Bi-LSTM as the feature extraction part of model. MKL-SVM is used to replace the Softmax function in Bi-LSTM to classify the data after feature extraction. By adding reverse time sequence, the Bi-LSTM-MKL-SVM network model can consider more time sequence factors, achieve better prediction effects for time sequence data, and carry out more effective feature extraction for time sequence data. It has better classification ability for heterogeneous data and reduces the time used for reference adjustment depending on experience. The feature extraction ability and generalization ability of the Bi-LSTM-MKL-SVM network model were verified by the confusion matrix, PR curve, and accuracy comparison. The accuracy of the Bi-LSTM-MKL-SVM network model was higher. The Bi-LSTM-MKL-SVM network model can improve the accuracy of transformer fault diagnosis and reduce the misjudgment of faults in the operation of transformers, thus reducing the economic losses caused by transformer faults.

The model proposed in this paper has certain limitations on the analysis of gas data. Different fault types pay different attention to gas data. More accurate selection of the weight of each gas in different faults can effectively improve the accuracy of classification. Transformer failure is closely related to environmental factors, such as ambient temperature and humidity. Considering more influencing factors to improve the universality of the model is one of the future research directions.

## ACKNOWLEDGMENT

This work was supported by the Scientific Research Project of the Liaoning Provincial Department of Education (No. LJKZ0398) and the Scientific Research Foundation of Liaoning Petrochemical University (No. 2017XJJ-044). The authors also would thank reviewers for improving this article.

## REFERENCES

- [1] Y. H. Wang, and Z. Z. Wang, "Transformer fault identification method based on RFRFE and ISSA-XG Boost," *Journal of Electronic Measurement and Instrumentation*, vol: 35, pp. 142-150, 2021.
- [2] H. Zhang, and S. X. Zhu, "Application of support vector machine classification in asynchronous motor fault diagnosis," *Journal of Suzhou University of Science and Technology Engineering and Technology edition*, vol: 32, pp. 70-74, 2019.
- [3] X. Zhang, H. Wang, Y. P. Wei, S. L. Wang, and Y. H. Su, "Fault diagnosis of power transformer based on SSA-PNN," *Industrial Instrumentation & Automation*, vol: 1, pp. 86-90, 2022.
- [4] Q. C. Fan, F. Yu, and M. Xuan, "Transformer fault diagnosis method based on improved whale optimization algorithm to optimize support vector machine," *Energy Reports*, vol: 7, pp. 856-866, 2021.
- [5] Y. H. Wu, X. B. Sun, Y. Zhang, X. J. Zhong, and L. Cheng, "A Transformer Fault Diagnosis Based on Improved Squirrel Search Algorithm and Support Vector Machine," *Journal of Physics: Conference Series*, vol: 2203, 012067, 2022.
- [6] B. Zeng, J. Guo, W. Q. Zhu, Z. H. Xiao, F. Yuan, and S.X. Huang, "A Transformer Fault Diagnosis Model Based On Hybrid Grey Wolf Optimizer and LS-SVM," *Energies*, vol: 12, pp. 1-18, 2019.
- [7] J. J. Miao, D. B. Li, H. J. Li, and P. Y. Liu, "Research on Furnace Slagging Prediction based on Improved Particle Swarm Optimization and Fuzzy Neural Network," *Journal of Engineering for Thermal Energy and Power*, vol: 37, pp. 104-114, 2022.
- [8] X. C. Gao, Q. Li, L. L. Wang, G. X. Du, and X. Ke, "Adaptive Convolutional Neural Network Based on Modified Genetic Algorithm," *Computer Technology and Development*, vol: 32, pp. 132-136+142, 2022.
- [9] I. B. M. Taha, S. Ibrahim, and D. E. A.Mansour, "Power Transformer Fault Diagnosis Based on DGA Using a Convolutional Neural Network with Noise in Measurements," *IEEE ACCESS*, vol: 9, pp. 111162-111170, 2021.
- [10] X. D. Fan, W. P. Fu, Z. L. Zhao, S. T. Zu, L. S. Zhang, and W. T. Hu, "Study on Fault Diagnosis of Oil-Immersed Transformer Based on Long-Short Term Memory Network," *Transformer*, vol: 58, pp. 27-32, 2021.
- [11] X. X. Wu, Y. G. He, J. J. Duan, H. Zhang, and Z. R. Zeng, "Bi-LSTM-based transformer fault diagnosis method based on DGA considering complex correlation characteristics of time sequence," *Electric Power Automation Equipment*, vol: 40, pp. 184-193, 2020.
- [12] O. A. Alharbi, "Deep Learning Approach Combining CNN and Bi-LSTM with SVM Classifier for Arabic Sentiment Analysis," *International Journal of Advanced Computer Science and Applications*, vol: 12, pp. 165-172, 2021.
- [13] K. Wang, J. Z. Li, and S. Q. Zhang, "New Features Derived from Dissolved Gas Analysis for Fault Diagnosis of Power Transformers," *Proceedings of the CSEE*, vol: 36, pp. 6570-6578+6625, 2016.

- [14] X.D.Pei, L.Luo, and S. Chen, "A Convolutional Neural Network Diagnosis Method for Dissolved Gas in Power Transformer Oil," *Journal of Liaoning Petrochemical University*, vol:40, pp.79-85, 2020.
- [15] K. Gregor, I. Danihelka, and A. Graves, et al. , "DRAW: a recurrent neural network for image generation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 2015, pp. 1462-1471.
- [16] S. X. Liu, S. Z. Gao, Y. Liu, J. Li, and Y. D. Cao, "Residual Life Prediction of AC Contactor Based on LSTM," *High Voltage Engineering*, 2022, pp. 1-11.
- [17] X. Chang, Y. B. Li, and M. Tian, et al. , "Reinforcement Learning Algorithm Based on One-dimensional Convolutional Recurrent Network," *Computer Measurement & Control*, vol: 30, pp. 258-265, 2022.
- [18] N. S. Kiruthika, and G. Thailambal, "Dynamic Light Weight Recommendation System for Social Networking Analysis Using a Hybrid LSTM-SVM Classifier Algorithm," *Optical Memory and Neural Networks*, vol: 31, pp. 59-75, 2022.
- [19] G. Z. Huang, W. J. Li, and Y. H. Deng, "Modeling of Performance Decay Prediction Based on LSTM-SVM Flexible Circuit Board Processing Roll," *Journal of Physics: Conference Series*, vol: 1828, pp. 012027, 2021.
- [20] K. Hasegawa, and K. Funatsu, "Non-linear modeling and chemical interpretation with aid of support vector machine and regression," *Current computer-aided drug design*, vol: 6, pp. 24-36, 2010.
- [21] L. S. Nie, F. Y. Chang, X. Z. Chang, C. Liu, Y. W. Jin, G. S. Liu, J. S. Fu, and X. S. Han, "A Novel Self-adaptive Multiple Kernel Learning Algorithm," *Journal of Jilin University(Science Edition)*, vol: 59, pp. 1212-1218, 2021.

# DataOps Lifecycle with a Case Study in Healthcare

Shaimaa Bahaa<sup>1\*</sup>, Atef Z.Ghalwash<sup>2</sup>, Hany Harb<sup>3</sup>

Computer Science Dept., Misr University for Science & Technology (MUST)  
Giza, Egypt<sup>1,3</sup>

Computer Science Dept., Helwan University  
Cairo, Egypt<sup>2</sup>

<https://orcid.org/0000-0002-3458-0395>

**Abstract**—The DataOps methodology has become a solution to many of the difficulties faced by data science and analytics projects. This research introduces a novel DataOps lifecycle along with a detailed description of each phase. The proposed cycle enhances the implementation of data science and analytics projects for achieving business value. As a proof of concept, the new cycle phases are applied in a healthcare case study using the UCI Heart Disease dataset. Two goals are achieved. First, a dataset reduction by features analytic in which the four most effective features are selected. Second, different machine learning algorithms are applied to the dataset. The recorded results show that using the four most effective features is comparable with using the full features (thirteen features), and both approaches show high accuracy and sensitivity. The average accuracy of the highest four features is 82.32%, and the thirteen features is 84.28%. That means that the selected four features affect the applications with 97.67% accuracy. Besides, the average sensitivity of the highest four features is 87.94%, while the thirteen features are 87.12%. The study shows an interesting and significant result that data modeling needn't be done for all data science projects which reduced the dataset.

**Keywords**—DataOps lifecycle; DataOps in machine learning; DataOps in healthcare; DataOps in data science; feature extraction; feature selection

## I. INTRODUCTION

The fields of data science, analytics, and machine learning are expanding at an incredible rate. Businesses are now searching for experts who can sort through the data goldmine and assist them in making quick, informed business decisions. Although today, organizations have a great opportunity to access data-driven tools and business intelligence software. Most organizations fail to make business value from their investments in data [1]. Thus, resulting from the lack of maturity in data science projects, most implementations are laptop-based research projects that never impact customers. In addition to, local applications that are not built to scale for production workflows, or high-cost IT projects. Therefore, selecting the method for implementing a data-driven project must be done carefully to help when maintaining or even adding a new feature(s).

The legacy architecture and tools that require special skills to use by data scientists have become bottlenecks for business stakeholders. These tools are costly resources, especially when producing unplanned data analysis. Machine Learning and Artificial Intelligence algorithms are just tips of the iceberg [2] for getting business and customer value from data.

Therefore, the operation of affecting data is the most crucial aim.

Data analytics is used in business to help organizations make better business decisions to meet and increase customer value. Data draws beneficial conclusions by collecting and organizing it; a data-driven process covering everything from data collection to analysis. DataOps has emerged to meet such requirements. *DataOps* is an emerging set of practices, processes, and technologies for building and enhancing data and analytics pipelines [3]. The term DataOps is a merge of data and operations which was first introduced by Lenny Liebmann in a 2014 blog post titled "3 reasons why DataOps are essential for big data success". The term wasn't popularized until Andy Palmer's 2015 blog post "From DevOps to DataOps". Since then, interest has grown when the term DataOps was included in Gartner's "Hype Cycle" for data management in 2018 [4]. As Agile has a manifesto [5] for its 4s principles. DataOps has its own manifesto [6] too, which consists of 18 principles, unlike Agile 12 principles. The DataOps manifesto has been published by Christopher Bergh, Gil Bengiat, and Eran Strod [4]. The DataOps manifesto principles have complemented the initiative that came out in 2018 called "The DataOps Philosophy".

The problem DataOps has come up for solving and minimizing analytics "cycle time" between the proposal of a new idea and the deployment of finished analytics. For example, many organizations require months of cycle time to deploy 20 lines of SQL. The long cycle times are the primary reason analytics projects fail [7]. This has led to discouraged and disappointed users and disturbing creativity. The factors that lengthen cycle time are *Poor Teamwork, Lack of Group Cooperation, Waiting for Systems, Waiting for Data Access, Over-Caution, Requiring Approvals, Inflexible Data Architecture, Process Bottlenecks, Technical Debt and Poor Quality*, which were mentioned in [8]. These obstacles pushed data experts to find an effective solution; therefore, DataOps came up. DataOps's goal for data science is to turn unprocessed data into a useful data science product. DataOps has provided utility to customers through a rapid, scalable, and repeatable process.

Data experts have given DataOps many definitions depending on their points of view. As a result, there have been several attempts to define the concept of DataOps. For example, Gartner [9] defined DataOps as a collaborative data management practice focused on improving the communication, integration, and automation of data flows

between data managers and data consumers across an organization. While Eckerson [10] Group defined DataOps as an engineering methodology and set of practices designed for the rapid, reliable, and repeatable delivery of production-ready data, operations-ready analytics, and data science models. DataKitchen [11] said that DataOps is a collection of technical practices, workflows, cultural norms, architectural patterns, and much more. However, the most appropriate definition adopted here is that DataOps is a methodology that applies Agile development, DevOps, and lean manufacturing principles [12], all together to data analytics development and operations where they are the intellectual heritage for DataOps.

Agile is an application of the theory of constraints to software development, in which smaller lot sizes decrease work-in-progress and increase overall manufacturing system throughput. DevOps is a natural result of applying lean principles, for example, eliminating waste, continuous improvement, and broad focus on application development and delivery. Lean manufacturing also contributes a relentless focus on quality using tools such as statistical process control, to data analytics [11]. Due to different DataOps definitions, trying to evaluate different solutions and determine whether they will help to achieve DataOps goals or not is a confusing matter. The authors in [13] introduced a [DataOps Vendors Landscape](#) which was organized by the six key capabilities required for DataOps success.

The major contributions of this paper are summarized in the following points:

- Introducing a novel approach for the DataOps lifecycle with a detailed description for each phase.
- The most effective features selection and extraction.
- Proving that data modeling is not necessary for all data science and analytics projects.
- Presenting a case study in healthcare as a proof of concept.
- Dataset reduction for the UCI Heart Disease dataset.
- A comparison between different machine learning algorithms that have applied to the dataset for both the highest four features and all (thirteen) features.

This paper is organized as it follows. Section II presents the related work. Section III introduces the proposed DataOps lifecycle. Experiments and recorded results for a case study in healthcare are shown in Section IV. Section V has the conclusion and future work.

## II. RELATED WORKS

This section will investigate both DataOps related works and the UCI Heart Disease dataset related works. The number of DataOps related works are quite a few because of being a new research field. The author in [14] has illustrated the broad character of DataOps and shown that it is not a particular method or tool. However, a collection of principles and a way of doing things on a cultural, organizational, and technological level. He differentiated between the exploration of DataOps as

a discipline, which includes methods, technologies, and concrete implementations, and the investigation of the business value of DataOps. While authors in [15] have defined DataOps as an application of DevOps to data, which means how effective data operations can be when DevOps concepts are applied to data for managing and deriving analytics. They also outlined the DataOps process and platform as well as the data challenges in the manufacturing and utilities industries. In [16], the authors said that DataOps is a new approach that aims to improve the quality and responsiveness of the data analytics lifecycle. In addition, they broadly organized dataOps into three steps: build, execute, and operate.

The lifecycle of a DataOps process has been illustrated in [17]. Besides, it illustrates the main collaborators in the DataOps process in charge of generating business value. In addition, they have gathered and highlighted good practices in DataOps reported in the academic literature, which serves as a starting point. While [18] defines DataOps as a method for accelerating the delivery of high-quality results through automation and orchestration of data life cycle phases. Furthermore, a case study in collaboration with Ericsson was conducted and introduced. They used the key phases of the data analysis methods to explore the key phases of the data besides checking their similarities to the popular DataOps approach. The common limitations of the above related works were either ambiguity of the DataOps lifecycle or the shortage of applications.

In [19], the model they proposed has four phases: first, data gathering that was the UCI Machine Learning dataset. Second, they used two methods for the features selection: Pearson's Correlation Heatmap where they selected 9 features and Chi Squared Test that selected 6 features. The third stage consists of K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Decision Tree (DT). After applying those algorithms, they have further used Stacking and Voting ensemble techniques for better results. Although, their model performed better when they have used Pearson's Correlation Heatmap selected features. Their model has some limitations that it has taken more time to generate outcomes.

While [20], they used Logistic Regression (LR), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), Support Vector Machines (SVM), Xtreme Gradient Boosting Machine (XGBM), Light Gradient Boosting Machine (LGBM), and K-Nearest Neighbors (KNN) for the prediction of the UCI Machine Learning dataset individually. Then multi-model ensembles were created (Ensemble 1 and Ensemble 2), which have far higher accuracies than individual models. The models with the best values of the evaluated parameters were gathered. In order to train and test the models on five distinct folds and to determine the optimal values for the hyperparameters in each of the implemented classification algorithms, fivefold cross validation and GridSearchCV were employed. They used all the UCI Machine Learning dataset features in addition to, they grouped models to reach their accuracy and that were their limitations.

Also [21], chose the well-known Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest

Neighbor (K-NN) algorithms for the UCI Machine Learning dataset prediction; also, for Hungarian cardiovascular disease dataset. The proposed system consisted of data acquisition, pre-processing, feature/attribute selection, classifications, and performance evaluation. The FCBF and mRMR were the feature selection algorithms. They used the info gain function selection method that's available on Weka for actual feature ranking. The top 10 features were ranked, which was a limitation.

Authors in [22] present three approaches. First Approach was without doing feature selection and outliers detection. The second approach was with doing feature selection and no outliers detection. The third approach was doing feature selection and also outliers detection. In all approaches they used Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), and XGBoost ML algorithms for the UCI Machine Learning dataset prediction. Their results they have been reached needn't to do three approaches. In addition, the feature they chose to be either important or not for heart disease predication were a limitation.

### III. PROPOSED DATAOPS LIFECYCLE

Data analytics projects are failing despite qualified people, powerful tools, and huge investments. Data scientists spend 75% [23] of their time massaging data and executing manual steps. Slow and error-prone development disappointed and frustrated data team members and stakeholders.

According to the adopted definition, to manage data in accordance with corporate objectives, DataOps combines DevOps and Agile approaches. For instance, DataOps would position data to make recommendations for better product marketing, converting more leads, if the goal was to increase lead conversion rate. While DevOps procedures are utilized for code optimization, product builds, and delivery, Agile techniques are employed for data governance and analytics development. DataOps uses statistical process control (SPC) to continuously monitor and verify the data analytics pipeline, much like lean manufacturing does. SPC increases data processing efficiency, improves data quality, and ensures that statistics are kept within reasonable bounds. SPC helps to notify data analysts right away in the event of an anomaly or error so they can respond.

This study proposes a novel DataOps lifecycle as shown in Fig. 1, along with a detailed description for each phase. The significance of this cycle is to investigate and highlight that data modeling is not necessary for all data science and analytics projects that reduce the dataset. The following subsections will illustrate each phase in detail.

#### A. Define Data Domain

A common mistake in data science projects is the confusion between defining a data area and a data domain. A data domain is a specific area in a large area. For example, if we have a data science project in healthcare, the area is healthcare, while the domain is heart disease. This is called the

first data domain definition. The second data domain definition is what disease in the heart disease area we're going to work with.

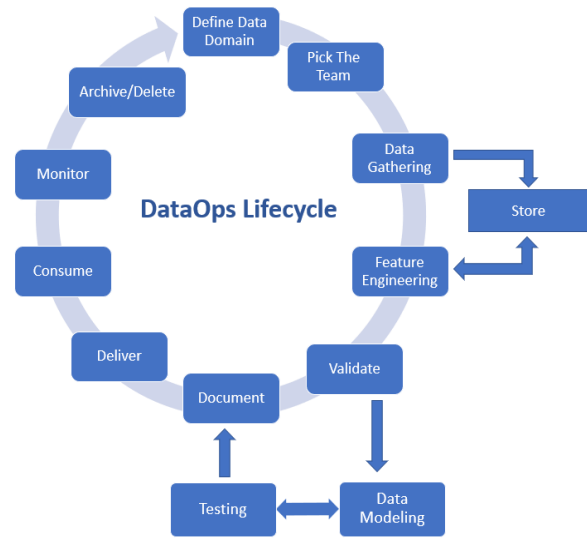


Fig. 1. Proposed DataOps lifecycle

#### B. Pick the Team

From this study point of view, this phase is the most critical phase that will determine the success of the coming phases or not. Data scientists have been working on different tools and outputting the results in different forms. This common problem adds another step to reorganize the output and merge all the results. Table I defines the criteria that must be met to approve each team member and the team as a whole.

Picking the team may differ from task to task. So, a smaller team would be selected by the unicorn for simpler problems. There are two main types of tasks, and each of them will be handled differently: First, system creation from scratch including data flow, data architecture, data schema, integration, etc. Second, solve smaller tasks from working with existing data or data that can be obtained. Each will require different handling methods and will be taken into consideration during the selection process.

#### C. Data Gathering

After everything was agreed upon, the team should start data gathering related to the domain. Data may be ready-made or performed from scratch. This phase should be done carefully as it may need to include any additional data, like performing surveys or scraping some data from the web. Then, the collected or performed data has to be stored.

#### D. Feature Engineering

Because not all data is important, this phase is concerned with obtaining and extracting the important data features. So, based on the data domain expert and, of course, the unicorn's opinion, the team tries to extract only the important features.

TABLE I. TEAM CRITERIA

Criteria	Criteria Details	
Members and Tasks	1. Data Unicorn	Data scientist who has the knowledge for all aspects of data science projects as, data engineering, statistical analysis, business analysis, ML, programming, or visualization. He must approve all the other members and he manages the whole project.
	2. An expert of domain knowledge	Any data project must have an expert in domain knowledge (with no experience in SWD) to explain all domain details and tricks the team want to know, i.e., educational expert in education project and medical expert in medical project.
	3. Data Analyst	Analyzes data (i.e., Visualizations: Charts, Graphs, Dashboards, Tables, Reports).
	4. Data Engineer	Develop, constructs, tests & maintain complete data architecture (i.e., Schema design, Data lakes).
	5. Data Scientist	Analyzes and interprets complex data in addition, he is a data wrangler who organizes (big) data.
	6. DataOps Engineer	He creates the mechanisms for workflow, manages cycle time and optimizes the quality.
	7. AI Engineer	Create end-to-end applications that include the data model(s).
	8. Operations Engineer	Deploying the applications into production environments and support service-level agreement.
Common Knowledge	They all must know source control, containerization, clean code, design pattern, security (to some level).	
Tools	All members must work with the same tool, or different tools that their output format is the same (provided that they will unify this later).	
Agreement	They must respect the data security and environment ethics.	

The next phase will be stepped to get the insights. The featured data may need to be stored or to store the featured criteria that have been adopted for similar data.

#### E. Store (Optional)

After collecting the data, it must be stored initially. Then, it might be restored or stored again in some other format after the next phase is done. For the sake of memory, the team might choose to store the data again or override the stored data based on its importance and size.

#### F. Validate

In this phase, the team seeks to gather all the collected insights. The task goal will be tried to reach based on the provided information. If they did, they wouldn't have to do the data modelling phase as results might be easily concluded only from the insights relying on the experience of the team. Data modelling leads to a lot of time, money, and resources; therefore, avoiding it on small tasks increases the speed. So, based on the problem type and the data, the team may also select a collection of algorithms and combine them together to obtain the best result for reaching the task goal. Thus, the team must successfully manage this phase in order to avoid the next if possible.

#### G. Data Modeling

If this phase is applied in the proposed cycle, it means that the team has tried every possible scenario to understand the data. Nevertheless, it was rather too complicated or critical for data modelling to be avoided. The team then must select the most perfect and appropriate algorithm for data modeling.

#### H. Testing

After reaching the goal, which of course, may differ from one task to another, in which data tasks need to be analyzed, visualized, modeled, processed, stored, or some of them, or all of them at the same time. The team must test whether the

results match the original goal or not. It must be checked step by step, there is no overfitting, underfitting, or any other data-related problems. The testing phase also must be reproducible. As the data modelling phase may be required to resolve a tested issue(s).

#### I. Document

All the work that has been done must be well described and documented. The documentation must be self-explanatory so that any member of any department can easily understand what has been done. Documentation may also involve technical writing for developers to complete from where the team stopped. Besides, if new data is created, a new data schema must be written to fairly describe the new changes. In addition to, the reasons why they needed to change the original data must be defined.

#### J. Deliver

The task delivery could be a tricky process. If the task contains new data, it must be put in the right place without any contradiction with the original data or any other data. If it contains reports, it must be very readable with visualizations to make the image clearer. Finally, if it has a new model delivered, it must be uploaded to the place where it will be used. So, an API could be created or maybe a model with a specific format. Therefore, the delivery must be teamwork to get the results in the right format to be used and easily deployed in production.

#### K. Consume

At an enterprise level, the published model(s) can be reused to derive various analytics required for business. "Recommend" solutions for understanding consumer preferences can be applied to a host of product lines. The tested and deployed solutions can be used with similar data sets to solve similar problems. This way not only does the enterprise save time by quickly applying proven



methodologies, but it also ensures the building of robust solutions through the continuous evolution process.

#### L. Monitor

Tasks delivery and doing investigations is not the end of the work. All the tasks delivered must be monitored to see whether the team succeeded in doing reliable work in the long term or not. For example, the team may have been assigned to work on a part that is rarely used. Besides, the team may have focused on the wrong features at first, and it was so obvious. Furthermore, monitoring is essential to predict and avoid an immediate system failure or even a small failure. Expect any changes before they happen. Detecting any possibility of performance reduction and being prepared for what changes may come may be enough. Thus, monitoring is essential for further improvement.

#### M. Archive/Delete

By the end, the organization might decide to cull out unwanted data or archive it to optimize resources and size management as well. Periodic data audits should be carried out to ensure production systems use fewer resources, running more efficiently and reducing storage costs overall. Data archiving plans have to be made for easy retrieval and more cost-effective information storage. Furthermore, irrelevant data needs to be purged.

## VI. EXPERIMENTS AND RESULTS

This section introduces a healthcare case study for the proposed DataOps lifecycle. Healthcare was chosen as it is the most sensitive data in which the results affect human life. According to WHO [24], cardiovascular diseases are the leading cause of death globally. It takes an estimated 17.9 million human lives each year. So, the chosen dataset was the UCI Machine Learning Heart Disease dataset [25]. This dataset contains 76 attributes, but all published experiments refer to using a subset of 14 (13 feature and target column) of them.

The "goal" which is the target, refers to the presence of heart disease in the patient. It is an integer valued between 0 (no presence) and 1 (presence). The other 13 features are:

- 1) **age**, that stores the age of the patient in years.
- 2) **sex**, where 1 is for males and 0 for females.
- 3) **cp**, chest pain type in which 1 = typical angina, 2 = atypical angina, 3 = non-anginal pain, and 4 = asymptomatic.
- 4) **trestbps**, which is resting blood pressure that was measured in mm Hg on admission to the hospital.
- 5) **chol**, serum cholesterol measured in mg/dl.
- 6) **fbs**, is fasting blood sugar > 120 mg/dl if 1 then true, while 0 = false.
- 7) **restecg**, resting electrocardiographic results that has three values 0 indicates normal, while 1 indicates having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) and 2 showing probable or definite left ventricular hypertrophy by Estes' criteria.
- 8) **thalach**, includes the maximum heart rate value achieved.

9) **exang**, exercise induced angina where 1 means yes and 0 means no.

10) **oldpeak**, that equals ST depression induced by exercise relative to rest.

11) **slope**, which is the slope of the peak exercise ST segment that has 3 values, if 1: upsloping, 2: flat, and 3: downsloping.

12) **ca**, includes a number of major vessels (0-3) coloured by flourosopy.

13) **thal**, where 3 = normal, 6 = fixed defect, and 7 = reversible defect.

The proposed DataOps lifecycle shown in Fig. 1 was applied as follows:

#### A. Define Data Domain

As illustrated in Section 3.1, the data area in this case study is healthcare. But the data domain is cardiovascular diseases.

#### B. Pick the Team

For this stage, as explained in Section 3.2. Choosing the team may differ from task to task. Therefore, in this case study, two main members from Table I must be present, in which they are an unicorn and a domain knowledge expert at least.

#### C. Data Gathering

After starting data gathering related to the domain. The UCI Machine Learning Heart Disease dataset has been chosen.

#### D. Feature Engineering

This phase has been done using Python 3 and the Jupyter Notebook IDE. Feature engineering has been done as follows:

- **Step 1:** Read/Load dataset.
- **Step 2:** Get the dataset information. The dataset information has displayed features(columns) name, each feature datatype, datatype, and total number of rows and columns.
- **Step 3:** Checking for null values.
- **Step 4:** Checking for duplicate.
- **Step 5:** Remove duplicate.
- **Step 6:** Generate Correlation heatmap. Where correlation heatmap is an essential step for data analysis, exploratory information in a visually appealing way. The value of the coefficient of correlation can take any value from -1 to 1 [26]. When the value is 1, it's a direct correlation between the two variables. That means when one variable increases, the opposite variable also increases. While if the value is -1, it's an indirect correlation between two variables, in which when one variable increases, the opposite variable decreases. Therefore, when 0 value, there's no correlation between two variables as the variables

change in a random manner with reference to one another.

E. Store

This stage may be visited many times. In this case, it has been visited twice. One for storing the gathered data and one after finishing the feature engineering stage.

F. Validate

In this phase, more data analysis has been applied to decide whether the data modelling phase is needed or not. Also, validate phase has been done using Python 3 and the Jupyter Notebook IDE. It may be done using author tools as Microsoft Excel or Power BI.

- **Step 1:** Display dataset description. A dataset description provides the following information for each feature: five number summary (minimum value, 25%, 50%, 75%, maximum value) in addition to mean, standard deviation, and count. For example, *age* feature five number summary (29, 47, 55, 61, 77 respectively). This means that minimum age in the dataset is 29, first quarter of age is 47, second quarter of age is 55, third quarter of age is 61, and maximum is 77 years. In addition to mean equals 54, with standard deviation 9, and count equals 303.
- **Step 2:** Generate correlation between target column and each feature in descending order. After tagging the absolute correlation values. This step gives the most effective features that correlate with target column. The most four effective features were *exang*, *cp*, *oldpeak*, and *thalach* as shown in Fig. 2.
- **Step 3:** Exploratory Data Analysis for each feature. This step gives more information about each feature. In addition, it represents heart disease root cause value for each feature. Table II illustrates each feature root cause value for having heart disease. The features have been arranged in descending order as same as in Fig. 2.

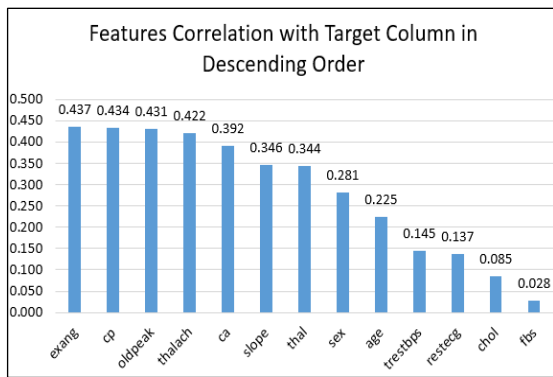


Fig. 2. Features correlation in descending order

Thus, in order to determine the most common causes of heart disease and what each cause truly affects, the validation

step is enough for doing that in addition to defining a set of values for heart prediction. While, if the task is the prediction of heart disease or not, data modelling needs to be done.

G. Data Modeling

For the heart disease prediction. This step has done to prove and support the proposed cycle. Two models have been developed. The same python code has been applied for both. One with all the features and the other with only the highest four features (*exang*, *cp*, *oldpeak*, and *thalach*) that highly affect the target. The two models were simply implemented. Using the following sklearn classifiers: Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF). In addition to that, XGBoost and Neural Network models with only one hidden layer and just 300 epochs.

H. Testing

Two test cases have been done. Where accuracy, f1 score, sensitivity, and specificity have been used in evaluation. According to [35,36], accuracy is the number of correctly classified data samples over the total number of data samples.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{1}$$

An F1 score is a measure of a test's accuracy.

$$F1\ Score = \frac{2TP}{2TP+FP+FN} \tag{2}$$

Sensitivity is the accuracy of a test to properly identify patients with a disease. In other words, it is the number of true positives divided by the number of actual positives.

$$Sensitivity = \frac{TP}{TP+FN} \tag{3}$$

Specificity is the accuracy of a test to properly identify people without the disease. This means that is the number of true negatives divided by all actual negatives. The test is positive if the person has the disease and, therefore, the test is positive. While true negative means the person doesn't have the disease and therefore the test is negative. A false positive is when the person doesn't have the disease and therefore the test is positive. A false negative means the person has the disease and therefore the test is negative.

$$Specificity = \frac{TN}{TN+FP} \tag{4}$$

In Case 1, the results for both developed models, the models with the highest four features and the models with all features, have been recorded in Tables III and IV, respectively. In Table III, the highest four features model has achieved great accuracies especially, for the neural network that was developed only with one hidden layer and 300 epochs. The accuracy recorded was 87.32% against 86.89 for the developed model with all features.

TABLE II. EACH FEATURE ROOT CAUSE VALUE

Feature	Value	Comment
exang	0	Samples without exercise induced angina are much likely to have heart disease. The same result was mentioned in [27].
cp	3	Samples with non-anginal pain are much likely to have heart disease. The same result was mentioned in [28].
oldpeak	[0, 1]	ST depression induced by exercise relative to rest with values from 0 to 1 are much likely to have heart disease.
thalach	[140, 170]	Samples with heart rate value from 140 to 170 are much likely to have heart disease.
ca	0	Number of major vessels equals 0 is much likely to have heart disease.
slope	3	The down sloping samples of the peak exercise ST segment are much likely to have heart disease.
thal	6	Fixed defect thalassemia samples are much likely to have heart disease.
sex	1	Males are much likely to have heart disease which was also mentioned in [29].
age	>= 40	Age grater or equal than 40 years old samples are much likely to have heart disease as mentioned in [30].
trestbps	[140, 200]	Resting blood pressure value from 140 to 200 samples are much likely to have heart disease as mentioned in [31,32].
restecg	1	Resting electrocardiographic of value 1 indicates having ST-T wave abnormality is much likely to have heart disease.
chol	[200, 300]	Serum cholesterol samples value from 200 to 300 are much likely to have heart disease as also mentioned in [33].
fbs	0	Samples with fasting blood sugar less than 120 mg/dl are much likely to have heart disease as mentioned in [34].

TABLE III. MODELS WITH HIGHEST 4 FEATURES COMPARATIVE ANALYSIS

Classifier	Accuracy	F1 Score	Sensitivity	Specificity
LR	85.96	86.21	89.29	82.76
NB	82.46	83.87	92.86	72.41
SVM	84.21	85.25	92.86	75.86
K-NN	80	79.52	84.62	76.09
DT	77.19	77.97	82.14	72.4
RF	82.46	83.33	89.29	75.86
XGBoost	78.95	79.31	82.14	75.86
Neural Network	87.32	86.15	90.32	85.0

In addition, the accuracies of LR, SVM and XGBoost were also greater than the same classifiers accuracies in Table IV. This proves that highest four features (exang, cp, oldpeak, and thalach) really affected the target column. Moreover, the sensitivities of the highest four features model were either higher than (as LR, NB, SVM and XGBoost) or comparable with the sensitivities of all features model. Thus means, the highest four features model greatly classify samples with heart diseases.

In Case 2, the recorded results for machine learning classifiers in the model with the highest four features were comparable to the results of both P. Gupta et al.'s model (with 13 features) [20] and Bharti et al.'s model (with 13 features) [22]. Fig. 3 to 5 illustrates the accuracy, sensitivity, and specificity comparisons respectively. The highest four features model classifiers (LR, SVM, RF and XGBoost) achieved higher accuracies than the same classifiers in Bharti et al.'s model. In consideration, their results were recorded after applying three approaches. Also, KNN, RF and XGBoost accuracies in highest four features were higher than P. Gupta et al.'s model. Moreover, the average accuracy of the highest

four features model recorded 81.6% against 82.15% in P. Gupta et al.'s model and 80.88% in Bharti et al.'s model.

TABLE IV. MODELS WITH ALL (13) FEATURES COMPARATIVE ANALYSIS

Classifier	Accuracy	F1 Score	Sensitivity	Specificity
LR	85.25	86.96	88.23	81.48
NB	83.61	85.71	88.23	77.78
SVM	83.61	85.71	88.23	77.78
K-NN	84.21	86.67	90.69	75.76
DT	82.89	85.06	86.04	78.79
RF	90.16	91.18	91.18	88.89
XGBoost	77.63	80.0	79.07	75.76
Neural Network	86.89	88.0	85.29	88.88

The sensitivities, shown in Fig. 4, of all classifiers in the highest four features model were higher than both the sensitivities in both P. Gupta et al.'s model and in Bharti et al.'s model (except for KNN was 84.64% against 85%). This means that the highest four features model has classified the samples with heart diseases greatly better than both comparable models. While the specificities of P. Gupta et al.'s model classifiers were higher than the highest four features model which mean it has classified samples without heart diseases better.

#### I. Document

All the previous steps have to be documented step by step.

#### J. Deliver

After writing the documentation, it is time to deliver all that has been done to the operations team for being deployed in production.

### K. Consume

In this study, the introduced models were decided to be consumed.

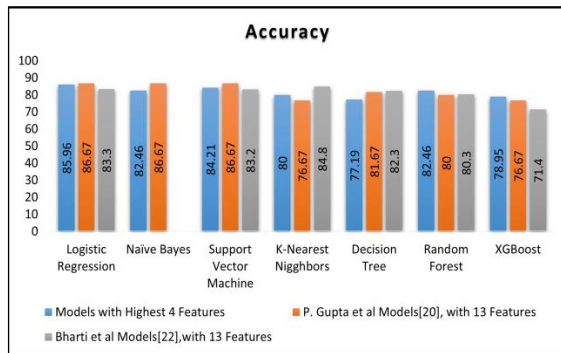


Fig. 3. Accuracy comparative analysis

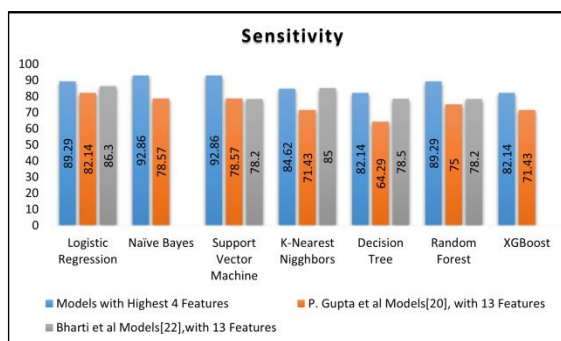


Fig. 4. Sensitivity comparative analysis

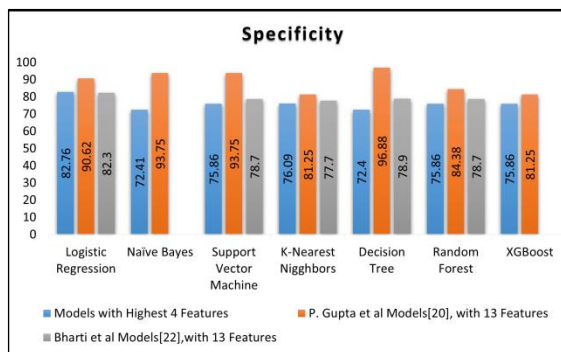


Fig. 5. Specificity comparative analysis

### L. Monitor

In real production, this phase must be done to observe models' behaviors.

### M. Archive/Delete

Choosing to do this phase will be an unicorn decision. In this study all the work has been archived.

## VII. CONCLUSION AND FUTURE WORK

This study proposed a novel DataOps lifecycle along with a detail for each phase that was applied to a healthcare case study. For this case, the UCI Machine Learning Heart Disease dataset has been used which has 13 features in addition to target column. The dataset has been analyzed without

modeling to find the best most effective features. This analysis came up with highest four effective features (exang, cp, oldpeak, and thalach) that affected the target column, indicating that data modeling is not necessary for all data science project which led to dataset reduction. Then, two models, with the same python code, have been developed for this dataset. The first included 13 features. The second one was with only the highest four (exang, cp, oldpeak, and thalach) features after removing duplicates in rows (20 rows). Two comparisons using accuracy, f1score, sensitivity, and specificity have been done.

Case 1 is the results of the developed model with 13 features and the model with only four features. The comparison showed that the highest four feature model produced higher accuracy and sensitivity, especially for neural networks, with 87.32% and 90.32%. Considering that the neural network model has been developed with only one hidden layer and just 300 epochs; also, the average accuracy and sensitivity of the highest four feature model was 82.32% and 87.94%, respectively.

Case 2, The results of machine learning classifiers in the four features model were comparable to both P. Gupta et al.'s model [20] and Bharti et al.'s [22] results. The comparison showed that DataOps gives great impact results when applied to machine learning model(s). The accuracy of classifiers in the highest four features model, with an average of 81.6%, was greater than the accuracy of Bharti et al., with an average of 80.88% and comparable with P. Gupta et al. with an average of 82.15%. The sensitivity of all classifiers in the highest 4 feature model, with an average of 87.6%, was greater than the sensitivity of both P. Gupta et al. with an average of 74.49% and Bharti et al., with an average of 81.24% while the specificities were comparable. In addition to that, Bharti et al.'s research mentioned that CP and thalach features [Fig. 5(c) & (e)] [22] were not important for heart disease unlike the proof from the proposed DataOps lifecycle.

For future work, the proposed DataOps lifecycle may apply to other fields such as the economy, education, or industry aside from applying it to deep learning model(s).

### NOTES

For any additional information or more explanation about either the proposed cycle or the code, feel free to contact co-author.

### REFERENCES

- [1] R. Bean, "Why Is It So Hard to Become a Data-Driven Company?" Harvard Business Review, Feb 2021. [Online]. Available: <https://hbr.org/2021/02/why-is-it-so-hard-to-become-a-data-driven-company>. [Accessed November 2021].
- [2] B. Chadha and S. Juwe, Agile Machine Learning with DataRobot, Packet Publishing, 2021, pp. 4-7.
- [3] W. Eckerson and J. Earth, "DataOps: Industrializing Data and Analytics," Eckerson Group, 2019.
- [4] H. Atwal, Practical DataOps: Delivering Agile Data Science at Scale, Isleworth, UK: Springer, 2020, p. 20.
- [5] R. C. M. e. a. Kent Beck, 2001. [Online]. Available: <https://agilemanifesto.org/principles.html>. [Accessed November 2021].
- [6] G. B. E. S. Christopher Bergh, "DataOps Principles," 2017. [Online]. Available: <https://dataopsmanifesto.org/en/>. [Accessed December 2021].

- [7] DataKitchen, "Minimizing Analytics Cycle Time with DataOps," 2017. [Online]. Available: <https://medium.com/data-ops/minimizing-analytics-cycle-time-with-dataops-b1a1b6e5ab22>. [Accessed December 2021].
- [8] "Enabling Design Thinking in," in *The DataOps Cookbook*, DataKitchen, 2019, pp. 116-117.
- [9] K. Graziano, G. Adams, W. Eckerson and M. Ferguson, "what is DataOps?," 2020. [Online]. Available: <https://www.truedataops.org/>. [Accessed December 2021].
- [10] D. Wells, "DataOps: More Than DevOps for Data Pipelines," 2019. [Online]. Available: <https://www.eckerson.com/articles/dataops-more-than-devops-for-data-pipelines>. [Accessed December 2021].
- [11] "What Is DataOps?," 2019. [Online]. Available: <https://datakitchen.io/what-is-dataops/>. [Accessed January 2022].
- [12] A. M. Francés, "Problems with your data? You need DataOps," 2021. [Online]. Available: <https://anjanadata.com/en/problems-with-your-data-you-need-dataops/>. [Accessed January 2022].
- [13] B. Pfeifferle, "The DataOps Vendor Landscape, 2021," 2021. [Online]. Available: [https://dataops.datakitchen.io/pf-eckerson-everything-you-need-to-know-about-dataops-solutions/pf-blog-dataops-vendor-landscape?utm\\_source=datakitchen&utm\\_medium=referral&utm\\_campaign=webinar\\_eckerson\\_dataops\\_solutions](https://dataops.datakitchen.io/pf-eckerson-everything-you-need-to-know-about-dataops-solutions/pf-blog-dataops-vendor-landscape?utm_source=datakitchen&utm_medium=referral&utm_campaign=webinar_eckerson_dataops_solutions). [Accessed January 2022].
- [14] J. Ereth, "DataOps – Towards a Definition," *LWDA*, pp. 104-112, 2018.
- [15] P. R. Sahoo and A. Premchand, "DataOps in Manufacturing and Utilities Industries," *International Journal of Applied Information Systems (IJ AIS)*, 2019.
- [16] A. Capizzi, S. Distefano and M. Mazzara, "From DevOps to DevDataOps: Data Management in DevOps processes," *arXiv[cs.SE]*, 2019.
- [17] M. Rodriguez, L. Jonat and M. Mazzara, "Good practices for the adoption of DataOps in the software industry," *Information Technologies, Telecommunications and Control Systems (ITTCS)*, 2020.
- [18] A. Raj, D. I. Mattos, J. Bosch, H. H. Olsson and A. Dakkak, "From Ad-Hoc Data Analytics to DataOps," in *International Conference on Software and Systems Process*, South Korea, 2020.
- [19] F. Rahman and M. A. Mahmood, "A Comprehensive Analysis of Most Relevant Features Causes Heart Disease Using Machine Learning Algorithms," in *Proceedings of the International Conference on Big Data, IoT, and Machine Learning*, 2022.
- [20] P. Gupta, S. Mala, A. Shankar and V. S. Asirvadam, "Heart Disease Detection Scheme Using a New Ensemble Classifier," in *Advances in Data and Information Sciences*, 2022.
- [21] Z. Alom, M. A. Azim, Z. Aung, M. Khushi, J. Car and M. A. Moni, "Early Stage Detection of Heart Failure Using Machine Learning Techniques," in *Proceedings of the International Conference on Big Data, IoT, and Machine Learning*, 2022.
- [22] R. Bharti, A. Khamparia, M. Shabaz and G. Dhiman, "Prediction of Heart Disease Using a Combination of Machine," *Hindawi*, 2021.
- [23] "What Data Scientists Really Need," in *The DataOps Cookbook*, 2019, pp. 126-127.
- [24] "cardiovascular diseases," 2021. [Online]. Available: [https://www.who.int/health-topics/cardiovascular-diseases#tab=tab\\_1](https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1). [Accessed February 2022].
- [25] "Heart Disease Data Set," 1988. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/heart+disease>. [Accessed December 2021].
- [26] B. Williams and S. Cremaschi, "Data-Driven Model Development for Cardiomyocyte Production Experimental Failure Prediction," in *Computer Aided Chemical Engineering*, Elsevier, 2020, pp. 1639-1644.
- [27] L. Anderson, A. M. Dewhirst, J. He, M. Gandhi, R. S. Taylor and L. Long, "Exercise-based cardiac rehabilitation for patients with stable angina," *Cochrane Database Syst Rev*, 2017.
- [28] J. Constant, "The Diagnosis of Nonanginal Chest Pain," *The Keio Journal of Medicine*, vol. 39, no. 3, 1990.
- [29] S. H. Bots, S. A. E. Peters and M. Woodward, "Sex differences in coronary heart disease and stroke mortality: a global assessment of the effect of ageing between 1980 and 2010," *BMJ Global Health*, 2017.
- [30] J. L. Rodgers, J. Jones, S. I. Bolleddu, S. Vanthenapalli, L. E. Rodgers, K. Shah, K. Karia and S. K. Panguluri, "Cardiovascular Risks Associated with Gender and Aging," *J Cardiovasc Dev Dis*, vol. 6, 2019.
- [31] F. D. Fuchs and P. K. Whelton, "High Blood Pressure and Cardiovascular Disease," *Hypertension*, 2019.
- [32] E. Rapsomaniki, A. Timmis, J. George, M. Pujades-Rodriguez, A. D. Shah, S. Denaxas, I. R. White, M. J. Caulfield, J. E. Deanfield, L. Smeeth, B. Williams, A. Hingorani and H. Hemingway, "Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1.25 million people," *Elsevier Sponsored Documents*, 2014.
- [33] J. Roland, "What Is Serum Cholesterol and Why Is It Important?," *healthline*, 26 January 2017. [Online]. Available: <https://www.healthline.com/health/serum-cholesterol#prevention>. [Accessed 1 September 2022].
- [34] C. Park, E. Guallar, J. A. Linton, D.-C. Lee, Y. Jang, D. K. Son, E.-J. Han, S. J. Baek, Y. D. Yun, S. H. Jee and J. M. Samet, "Fasting Glucose Level and the Risk of Incident Atherosclerotic Cardiovascular Diseases," *Diabetes Care*, 2013.
- [35] S. Amelia, H. Roberta and T. Alison, "What are sensitivity and specificity?" *BMJ*, 2019.
- [36] H. N. B, "Confusion Matrix, Accuracy, Precision, Recall, F1 Score," Dec 2019. [Online]. Available: <https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd>. [Accessed Dec 2021].

# Evaluation of e-Service Quality Impacts Customer Satisfaction: One-Gate Integrated Service Application in Indonesian Weather Agency

Aji Prasetyo<sup>1</sup>, Deny Irawan<sup>2</sup>, Dana Indra Sensuse<sup>3</sup>, Sofian Lusa<sup>4</sup>, Prasetyo Adi Wibowo<sup>5</sup>, Alivia Yulfitri<sup>6</sup>  
Faculty of Computer Science, University of Indonesia, Jakarta, Indonesia

**Abstract**—Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) is the weather agency in Indonesia. It has One-Gate Integrated Service Application, also known as *Pelayanan Terpadu Satu Pintu* (PTSP) BMKG Application, which is a web-based e-commerce concept application. Its goal is to provide users with information and services related to Meteorology, Climatology, and Geophysics (MCG) by using information and communication technology. This is part of Indonesia's move toward e-government. Since January 2020, all MCG service and information activities through PTSP BMKG must be done using the application. With a questionnaire and multivariate analysis, this study aims to determine how the quality of service affects customer satisfaction with the PTSP BMKG application. Scientists use the E-S-Qual scale to prove that it works and is a good measure. The results of this study show that customer satisfaction is affected positively and significantly by efficiency, fulfillment, system availability, and privacy simultaneously. Partially, customer satisfaction with the PTSP BMKG application is affected positively and considerably by how well the application works and how well it meets customers' needs. This has implications for the evaluation that BMKG needs to do.

**Keywords**—e-Service quality; one-gate integrated service; e-government; multivariate analysis; Partial Least Square (PLS); Structural Equation Model (SEM)

## I. INTRODUCTION

The advancement of Information and Communication Technology (ICT) has provided convenience in social interaction, including in delivering public goods to the community [1]. The era of globalization, with the development of the ICT field, encourages a new paradigm, namely e-government, which is a reform of the implementation of government that refers to public information disclosure and gives responsibility to the government to provide information about government activities. Information technology is believed to have an essential role in changing the conventional way of doing government work to be more efficient and effective. Information technology is also one of the keys to supporting the implementation of e-government and good governance through increased government transparency and accountability.

Badan Meteorologi, Klimatologi dan Geofisika (BMKG), is tasked with carrying out government duties in meteorology, climatology, and geophysics [2][3]. It has been explained in the Presidential Regulation of the Republic of Indonesia Number 61 of 2008 that the need for data and information in the fields

of meteorology, climatology, and geophysics is necessary not only to support national development but also to protect society and minimize the impact of damage caused by disasters caused by nature. In all elements of the BMKG environment, in performing its duties, the BMKG must apply the principles of coordination, integration, and synchronization within the internal BMKG and in relations between central and regional government agencies [2][3].

The Public Service Act has mandated that every government institution establish an integrated service system formally and legally. Every institution has the right to have the authority to make public policies to provide services and increase participation in empowering society [4].

BMKG One-Gate Integrated Service Application, also known as PTSP BMKG Application is a web-based application intended to provide information and services related to MCG for users by utilizing information and communication technology so that it is hoped that public services can be achieved optimally in the digital transformation towards e-Government. PTSP BMKG application is currently focused on serving requests for data and information to customers, both individuals from the community and companies from Indonesia. Based on information from BMKG employees who participated in the development, this application is designed like a marketplace or e-commerce. e-Commerce refers to business conducted electronically or over the internet. Generally, it refers to individuals who use the internet to either purchase and sell goods or give paid services through payments and data flows [5][6]. In PTSP BMKG Application, Customers can choose the type of data, information, or services they want by entering the shopping cart, paying, and information can be downloaded on the application. The application can be accessed at <https://ptsp.bmkg.go.id/>.

Since January 2020, all MCG service and information activities through PTSP BMKG must be served using the application. Many complaints are related to services and information reported through the PTSP BMKG application. Of course, if not handled, the complaint will have a destructive impact on the government, especially the BMKG. Even worse is that it can generate distrust from users. In addition, based on Community Satisfaction Index Report in 2022 by the BMKG Database Center, the value of the "system" aspect experienced the highest decline. This component has an index value of 3.96 with a percentage of 79.12 percent in 2021, but it only reached a value of 3.81 with 76.29 percent in 2022. This can mean that

public satisfaction with the service system has decreased. Therefore, it is necessary to evaluate public services to direct actions that can positively impact the quality of services provided by the government [7].

Based on information from one of the BMKG Public Relations division employees, BMKG has surveyed customer satisfaction. Still, it has yet to include ICT aspects, especially web applications used by customers to obtain MCG services. Therefore, BMKG itself doesn't know whether there is an influence in terms of web quality on customer satisfaction in using the service.

e-Service Quality is a way to determine how a website can facilitate user activities that include product purchase and delivery transactions efficiently and effectively [8]. There are several measures for evaluating the quality of electronic services in the literature. The E-S-Qual measurement scale developed by Parasuraman et al. (2005) is well regarded and utilized by scientists to validate this scale's effectiveness [9][10][11]. To quantify the quality of electronic services more accurately from an e-commerce standpoint, this method successfully considers the concerns highlighted while maintaining the fundamental principles of E-S-Qual [12]. However, from previous studies [13][14][15], not all dimensions or variables of E-S-Qual can be accepted from hypothesis testing, so in this study, the authors tried to retest the dimensions contained in E-S-Qual for case the PTSP BMKG website as one of the government services, which has the concept of e-commerce. This study also aims to analyze the effect of service quality on customer satisfaction in using the PTSP BMKG application as one of the steps to evaluate BMKG services.

The study was organized into five sections. The first section is an introduction containing the research's background and objectives. The second section is a literature study containing previous research, explanations of fundamental theories, and conceptual framework. The third section describes the research methodology. The results and discussion are described in the fourth section. Finally, our work on this paper is concluded in the last section.

## II. STUDY OF LITERATURE

### A. Previous Research

We construct in this study a theoretical framework based on prior research on the measurement of E-S-Qual. Consider several dependent and independent variables, such as efficiency, fulfillment, system availability, and privacy, as a calculation of customer satisfaction.

Ahmed et al. (2020) used the E-S-Qual model to analyze customer satisfaction, the quality of electronic banking services, and the direct relationship between these two variables. The immediate research results showed that the E-S-Qual dimensions positively and significantly influenced customer satisfaction. The results of the indirect connection showed that customer satisfaction and E-S-Qual dimensions are mediated by perceived value and trust [13].

The E-S-Qual model has been the subject of more studies. The E-S-Qual technique is used by Dastane et al. (2018) to

research service aspects impacting customer satisfaction and loyalty. As a result, although efficiency, system availability, and responsiveness have no direct or indirect influence on customer happiness, fulfillment and security immediately impact it [14].

Research using E-S-Qual is also carried out in the public sector. This was done by Shafira (2021), who used the E-S-Qual methodology to research the quality of E-procurement services in Malaysia. In this study, three variables are E-S-Qual: efficiency, privacy, and system availability. The result is that efficiency, privacy, and system availability impact user satisfaction [15].

Based on these three studies, we need to retest all dimensions of the PTSP BMKG application used by BMKG to serve its users.

### B. E-S-Qual

The functional performance of a service when clients shop, buy and get it via electronic media is known as "e-service quality" [6][9]. In obtaining services and information from BMKG, customers are currently using electronic-based services, so electronic media presently facilitate the interaction between sellers and buyers in the form of sites specifically designed to carry out buying and selling transactions.

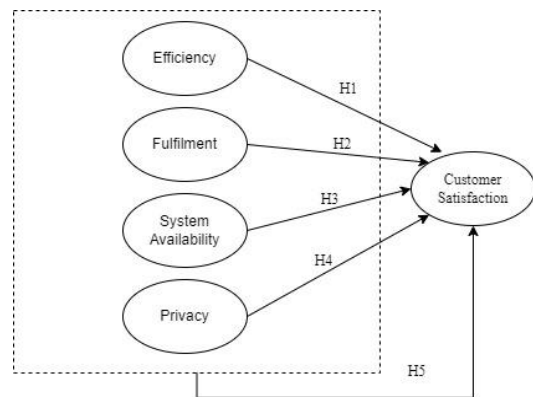


Fig. 1. Conceptual model

Several professionals have created electronic service quality measuring scales in various versions. One is E-S-Qual, which Parasuraman et al. created [9][11]. They argue that pre-existing measurement scale instruments still need to improve assessing the quality of electronic services because measurements cover only some stages in the interaction between users and services. Therefore, they developed two measurement scales, E-S-Qual, and E-RecS-Qual. There are four dimensions of E-S-Qual, including:

- Efficiency: The simplicity and quickness of utilizing the website.
- Fulfillment: How well the website delivers on its claims about order fulfillment and item availability.
- System Availability: the website's proper technical operation.
- Privacy: The website's level of security and privacy protection.

When an issue occurs throughout the service delivery process for customers, the E-RecS-Qual scale is used to gauge the quality of e-services [9]. In this study, the E-Recs-Qual scale was not included because in getting the service, customers can download data and information that has been ordered on the application or directly take it to the BMKG office.

C. Customer Satisfaction

According to Tjiptono, Customer satisfaction is the personal feeling of joy or disappointment generated by comparing perceived product performance (or results) with expected expectations [16]. In this study, we use several indicators of consumer satisfaction consisting of [17][18]:

- Expectation Conformity: Correspondence between consumer expectations and consumer perceptions.
- Returning Interest: Consumer willingness to revisit or reuse related services.
- Willingness to Recommend: Willingness of consumers to recommend recognized services to friends and family.

D. Conceptual Framework

Based on the theory in the literature study, Fig. 1 describes a research model that can be compiled that will be used to test the dimensions or variables of E-S-Qual on the PTSP BMKG application against its customer satisfaction.

From previous research, it is said that dependent and independent variables such as efficiency, fulfillment, system availability, and privacy have a positive effect as a calculation of customer satisfaction [13], so a hypothesis can be compiled in Table I.

TABLE I. HYPOTHESIS

Code	Description
H1	Efficiency positively and significantly affects customer satisfaction using PTSP BMKG Application.
H2	Fulfillment positively and significantly affects customer satisfaction using PTSP BMKG Application.
H3	System availability positively and significantly affects customer satisfaction using PTSP BMKG Application.
H4	Privacy positively and significantly affects customer satisfaction in using PTSP BMKG Application.
H5	Efficiency, Fulfillment, System Availability, and Privacy simultaneously positively and significantly impact Customer Satisfaction using PTSP BMKG Application.

Based on these five hypotheses, a questionnaire can be prepared. Customers of the PTSP BMKG Application will answer that by adjusting operational conditions.

TABLE II. STATEMENT OF QUESTIONNAIRE

Indicators (code)	Statement
Efficiency (X1)	<ol style="list-style-type: none"> <li>1. PTSP BMKG application makes it easy to explore what is needed related to Meteorology, Climatology, and Geophysics services and information [9][11].</li> <li>2. PTSP BMKG application makes it easy to quickly complete service transactions and information on Meteorology, Climatology, and Geophysics [9] [11].</li> <li>3. PTSP BMKG application can load pages quickly [9][11].</li> <li>4. All features in the PTSP BMKG application are well organized [9][11].</li> <li>5. The information in the PTSP BMKG application is well organized [9][11].</li> </ol>
Fulfillment (X2)	<ol style="list-style-type: none"> <li>1. PTSP BMKG application provides services according to what you need [9][11].</li> <li>2. PTSP BMKG application following the public service offer (completed within 14 working days) [9][11].</li> <li>3. PTSP BMKG application responds quickly to requests for services and information [9][11].</li> </ol>
System Availability (X3)	<ol style="list-style-type: none"> <li>1. PTSP BMKG application is always available when the user needs it [9][11].</li> <li>2. PTSP BMKG application rarely experiences errors [9][11].</li> </ol>
Privacy (X4)	<ol style="list-style-type: none"> <li>1. PTSP BMKG application doesn't disseminate customer personal information to other parties [9][11].</li> <li>2. PTSP BMKG application protects the personal information of its users [9][11].</li> </ol>
Customer Satisfaction (Y1)	<ol style="list-style-type: none"> <li>1. Customers are satisfied with the service of the PTSP BMKG application because the application service is in line with expectations. [17][18]</li> <li>2. Customers are satisfied with the PTSP BMKG application service, so they are interested in reusing it. [17][18]</li> <li>3. Customers are satisfied with PTSP BMKG application service to say positive things about the application to others [17][18]</li> </ol>

III. METHODOLOGY

This research is a quantitative study using primary data obtained from questionnaires. In Table II, the questionnaire is prepared based on the variables E-S-Qual and customer satisfaction. In each variable, statements are mapped based on the dimensions and indicators of each variable. The compiled questionnaire will be distributed to customers BMKG One-Gate Integrated Service Application. The stages of the study can be seen on the Fig. 2.

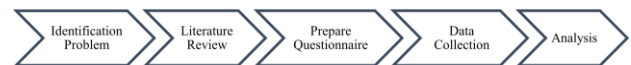


Fig. 2. The stages of study



### A. Data Collection

Data collection is carried out at BMKG by providing an online questionnaire that customers of the PTSP BMKG Application can fill out. The requirement to fill out the questionnaire is that the customer has transacted using the PTSP BMKG application. The population of this study is customers who have been registered and have used the PTSP BMKG application to obtain data and information from BMKG. However, in this study, using sampling, the number of respondents in this study was 72 respondents. A total of 85 respondents were distributed online questionnaires using email by attaching a google form link, assisted by the admin of the PTSP BMKG application; 13 customers did not fill out the questionnaire. The respondent's demographic profile of the data collection can be seen in Table III.

TABLE III. RESPONDENT'S DEMOGRAPHIC PROFILE

Items	Categories	Percentage
Gender	Male	68
	Female	32
Educational Background	High School	13
	The diploma I/ diploma III	7
	Bachelor's Degree	69
	Master's degree	11
Profession	Civil servant	17
	Private Employee	58
	Students	15
	Entrepreneur	4
	Others	6

The research questionnaire consists of several questions based on the E-S-Qual model and consumer satisfaction indicators. All items are measured using a Likert scale. The Likert scale was modified to clear doubts in respondents using a scale of four, including strongly agree, agree, disagree, and strongly disagree.

### B. Analysis

After collecting the questionnaire data, the next step is to process and analyze the data. The data obtained were processed using the *Smart PLS 4* application to test hypotheses and relationships between variables. This study uses a multivariate analysis method by testing the outer and inner models on PLS-SEM. A technique frequently used to process various variables is multivariate analysis [19]. The objective is to simultaneously look for these different variables' effects on an object. According to Hair et al. (2017b), by employing the PLS-SEM model, one can get beyond some of the drawbacks of covariance-based SEM, including data dispersion, limited sample sizes, complex models with numerous indicators, and model construction and testing [19] so that the results of the analysis can be used as the basis for evaluating e-service quality.

## IV. RESULTS AND DISCUSSION

### A. Assessing Reflective Measurement Models

According to Hair et al. [20], there are four stages in reflective measurement models: delivered *reflective indicator loadings, internal consistency, convergent validity, and discriminant validity*.

TABLE IV. OUTER LOADINGS

Code	Outer Loading	Code	Outer Loading	Code	Outer Loading
X1-1	0.822	X2-1	0.805	X4-1	0.902
X1-2	0.746	X2-2	0.784	X4-2	0.937
X1-3	0.793	X2-3	0.909	Y1-1	0.936
X1-4	0.872	X3-1	0.946	Y1-2	0.922
X1-5	0.855	X3-2	0.839	Y1-3	0.860

Based on Table IV, all outer loading on each indicator has a value above 0.708. More than half of the indicator's variability can be accounted for by the construct, which also ensures a sufficient level of dependability [20]. All indicators can proceed to the next stage.

Internal consistency reliability is the extent to which indicators measuring the same structure are related [19]. When evaluating internal consistency reliability, Joreskog's (1971) composite reliability should almost always be used [21][20]. The range of reliability values from 0.6 to 0.7 is considered "acceptable in research exploration," but the range of reliability values from 0.7 to 0.9 is classified as "satisfactory to good." However, results that are either equal to or higher than 0.95 are unsatisfactory because it is considered that these values are excessive, which can reduce the validity of the construct. Based on Table V, the composite reliability value on the fulfillment construct is the lowest, with a value of 0.873, and the highest in the system availability construct, with a value of 0.933. Therefore, all the construct validity is well accepted and satisfactory.

TABLE V. CONSTRUCT RELIABILITY AND VALIDITY

Indicators	Cronbach's alpha	Composite reliability	The average variance extracted (AVE)
Customer Satisfaction	0.891	0.933	0.822
Efficiency	0.877	0.910	0.671
Fulfillment	0.780	0.873	0.697
System Availability	0.763	0.888	0.799
Privacy	0.819	0.916	0.846

Convergent validity means a set of indicators explained by the construct. Representations can be demonstrated through one-dimensionality that can be expressed using Average Variance Extracted (AVE) values. An adequate convergent validity is defined as a construct's ability to account for more than 50 percent of the variance among its indicators [9][22]. In Table V, the AVE values for all variables are greater than 0.5; therefore, it can be concluded that all the indicators developed for this study demonstrate strong convergent validity [23].

Researchers employ discriminant validity to ensure that each construct's ideas are distinct. The purpose of validity testing is to determine how precisely a measuring instrument performs its measurement function [24]—calculating the discriminant validity of this study using the heterotrait-monotrait (HTMT) ratio. To evaluate the validity of the discriminants, Henseler, Ringle, and Sarstedt (2015) propose an alternative method based on the multitrait-multimethod matrix: HTMT [25][26]. Structural models with similar constructs should have a threshold value of 0.90 [20]. Based on Table VI, all indicators have a value of < 0.90; it can be interpreted that the matter is reasonable to describe the discriminant validity of each indicator used in this study.

TABLE VI. HETERO TRAIT-MONOTRAIT RATIO OF CORRELATIONS (HTMT)

Indicator	Y1	X1	X2	X4	X3
Y1					
X1	0.783				
X2	0.880	0.898			
X4	0.504	0.585	0.543		
X3	0.540	0.671	0.703	0.571	

**B. Assessing Structural Model**

The structural model should be checked for collinearity first, as stated by Hair et al. (2017). Examine the structural model's VIF values for each predictor construct [27]. In this study, no collinearity problems were found, which can be seen in Table VII. Collinearity values are still between 0.5 and 3.0. This follows what Cenfetelli & Bassellier (2009) said: It is possible to examine the VIF value, and if it is less than 3.0, multicollinearity is probably not a concern [28]. Testing bivariate correlations between construct scores is an alternate strategy. When the bivariate correlation is greater than 0.50, multicollinearity may impact the path coefficient's magnitude and/or direction [27].

TABLE VII. VIF VALUE

Indicators	Customer Satisfaction
Efficiency	2.572
Fulfillment	2.361
Privacy	1.374
System Availability	1.691

The next step is the significance and relevance of the path coefficients [17][22]. The results of the coefficient path test for direct effect can be seen in Fig. 3. The route coefficients are standardized numbers with a possible range of +1 to -1; however, they only sometimes get close to +1 or -1. This is especially true for large models with many separate structural elements [22]. This study found that efficiency, fulfillment, and system availability have a positive relationship with customer satisfaction, while system availability has a negative association with customer availability.

The following assessment recommended is multiple regression models or the coefficient of determination [27][22]. The coefficient of determination quantifies all endogenous constructions' in-sample prediction accuracy. Because the prediction only estimates the predictive power for the sample data used in the findings, R<sup>2</sup> should not be generalized to the population (Rigdon, 2012; Sarstedt et al., 2014) [22]. The results of the R<sup>2</sup> test showed a value of 59.6%. This can mean that 59.6% of customer satisfaction is due to efficiency, fulfillment, system availability, and privacy.

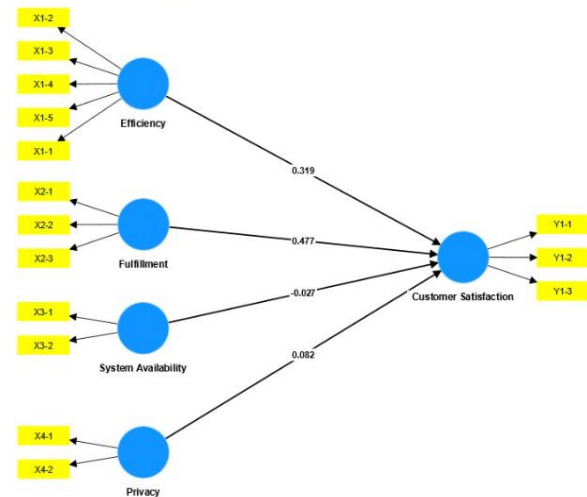


Fig. 3. Path coefficient

**C. Hypothesis Testing**

The first for hypothesis testing is the assessment of the T-statistics or P-value. The result can be seen in the table for the T-statistics is more than 1.96 ( $\alpha = 0,05$ ) [19] and for P-value less than 0.05 [20]. Based on the T-statistics and P-value in Table VIII, the first hypothesis (H1) is accepted since the score of the T-statistics is 2.086, more than 1.96, and the score of P-value is 0.019, less than 0.05. So that the first hypothesis, Efficiency positively and significantly affects customer satisfaction using the PTSP BMKG Application, is acceptable. Where Efficiency positively and significantly affects customer satisfaction is 31.9%. The effect of efficiency is 31,9% based on a value in the original sample.

The second hypothesis (H2) is validated by the T-statistics shown in Table VIII because the T-statistics score is 3,181, which is more than 1.96, and the P-value is 0.001, which is less than 0.05. So that the second hypothesis is that Fulfillment positively and significantly affects customer satisfaction using PTSP BMKG Application. The second hypothesis is

acceptable. Where Fulfillment positively and significantly affects customer satisfaction is 47.7%. The effect of Fulfillment is 47.7% based on a value in the original sample.

The T-statistics reject the third hypothesis (H3) in Table VIII because its score is 0,211, less than 1.96, and its P-value is 0.416, which is more than 0.05. So, system availability positively and significantly affects customer satisfaction using the PTSP BMKG Application is rejected. Based on the t-statistics, p-value, and original sample, System availability to customer satisfaction is negative and does not significantly affect customer satisfaction.

The fourth hypothesis (H4) is shown to be rejected by the T-statistics in Table VIII because the T-statistics score is 0.767, less than 1.96, and the P-value score is 0.222, more than 0.05. So, privacy positively and significantly affects customer satisfaction in using the PTSP BMKG Application is rejected. Based on the t-statistics, p-value, and original sample, privacy positively impacts customer satisfaction but does not significantly affect customer satisfaction.

TABLE VIII. ORIGINAL SAMPLE, T-STATISTICS, P-VALUES

Code	Original Sample	T-statistics	P-values	Conclusion
H1	0.319	2.082	0.019	Accepted
H2	0.477	3.181	0.001	Accepted
H3	-0.027	0.211	0.416	Rejected
H4	0.082	0.767	0.222	Rejected

To test the fifth hypothesis, we use simultaneous F testing. Simultaneous F tests are intended to find whether concurrently independent variables (simultaneously) impact dependent variables. The F test was performed to prove the impact of all free variables simultaneously on bound variables [24]. The test results found that the F value was 24.9 or greater than the F table, which were 2.51. This follows the fifth hypothesis (H5) that Efficiency, Fulfillment, System Availability, and Privacy simultaneously positively and significantly impact Customer Satisfaction using the PTSP BMKG application. Hence, the results of the fifth hypothesis are acceptable.

TABLE IX. F-SQUARE

Direct Relationship	F-Square	Impact
Efficiency → Customer Satisfaction	0.098	Small
Fulfillment → Customer Satisfaction	0.238	Medium
Privacy → Customer Satisfaction	0.012	-
System Availability → Customer Satisfaction	0.001	-

In this study, the F-square value is used to measure the extent to which several aspects of the PTSP BMKG application, including efficiency, fulfillment, privacy, and system availability, influence customer satisfaction. According to Hair et al. [29], if the value of f-square is more significant than or equal to 0.02 but is less than 0.15, then the effect has a minor influence and falls into the category of "small.". If the f-square value is greater than or equal to 0.15 and less than 0.35,

it belongs to the medium impact category. If the value of f-square is greater than or equal to 0.35, then it belongs to the category of a significant impact. Based on Table IX, the efficiency indicator has an f-square of 0.098, meaning that the efficiency indicator has a negligible effect on customer satisfaction using the PTSP BMKG application. At the same time, the fulfillment indicator has an f-square of 0.238. This can mean that the fulfillment indicator moderately impacts customer satisfaction using the application. The privacy and system availability indicators have a very small f-square value, so the two indicators have no impact on customer satisfaction using the PTSP BMKG application.

#### D. Implication of Study

This study has various implications for problems relating to the e-service practice paradigm, which affects consumer satisfaction in government apps utilized for public services, especially in the PTSP BMKG application, which is one of the BMKG services. BMKG needs to improve the quality of its application so that users can be more satisfied in obtaining services and information through PTSP, including in the dimension of efficiency, which focuses on ease of transactions, web speed, features, and information that is better laid out. Not only that, but the most crucial focus to be noticed and improved by BMKG is on the fulfillment dimension, where the speed of response to customer requests and submissions to get services is at most 14 working days.

This research can be used as a reference that demonstrates that E-S-Qual has developed into an efficient scale in government services with a concept like e-commerce, in which a website can help users with their activities, such as product purchases, and obtain transactions efficiently and effectively. Based on the theory explained by Parasuraman et al. (2005), efficiency and fulfillment are the most critical aspects of the quality of website services [6]. This is in line with the results of this study that these two dimensions have an impact on customer satisfaction in using applications to obtain data and information services from BMKG. The efficiency indicator itself is related to the design of the web application interface used by customers to get the information needed. Fulfillment indicators related to how all customer needs can be available and met quickly, as stated on the PTSP BMKG web application. Customers PTSP will feel more satisfied if the service can be obtained on time or faster than the promised time.

Wolfenbarger and Gilly (2003) claim that consumers who often use online applications to transact may not find privacy to be as crucial as they formerly did since experience may diminish worries about website security [30][9][14]. This is also in line with the results of this study that privacy does not significantly influence customer satisfaction who have transacted through the PTSP BMKG web application in obtaining data and information services from BMKG. Even so, BMKG must still maintain customer privacy, which has been regulated by the government so as not to cause leakage of customer data using the PTSP BMKG application.

The system availability aspect is an essential attribute in measuring overall customer satisfaction. However, BMKG does not fully control this dimension because the device and

internet used can affect this dimension [9]. System availability does not affect customer satisfaction with the PTSP BMKG application. It could be because customers using the application are fine with this dimension. It could also be due to the number of respondents, as said by Dastane et al. (2018) [14]. Even so, BMKG must still inform customers when the website is being maintained. BMKG must think of the right time to maintain the website PTSP BMKG so customers are not bothered when they want to use the application.

## V. CONCLUSION

BMKG Indonesia has changed how it serves the community conventionally to be more efficient and effective since 2020 by presenting the PTSP BMKG application. To improve service quality in obtaining customer satisfaction, it is necessary to evaluate the PTSP BMKG application. The research made use of the E-S-Qual instrument. Considering the outcomes of data analysis from 72 respondents of application users that have been carried out in this study, it can be concluded that the results of statistical calculations show Efficiency, Fulfillment, System Availability, and Privacy simultaneously has a positive and significant impact on customer satisfaction using the PTSP BMKG application.

But partially, we can conclude that the ease of users in searching to find what services and information are needed, the ease of completing transactions, the speed of loading web pages, and the features and availability of information that is better organized on the PTSP BMKG application have an impact on customer satisfaction in using the application. This is shown in the results of the H1 hypothesis test received, where efficiency positively and significantly affects customer satisfaction in using the PTSP BMKG application, so BMKG must be concerned about efficiency indicators so that customers can be more satisfied with the MCG data and information services obtained from the PTSP BMKG Application.

Moreover, the suitability of MCG's service and information needs and the speed of response to requests and submissions to obtain services and information will have a moderate impact on customer satisfaction in using the application. This is shown in the results of the H2 hypothesis test received, where fulfillment positively and significantly affects customer satisfaction in using the PTSP BMKG application. Customers will be happier if the service is provided on time or earlier than promised. Hence, BMKG needs to evaluate fulfillment indicators to be able to increase customer satisfaction with the PTSP BMKG application.

### A. Limitation of Study

This paper is focused on PTSP BMKG Application, which serves customers to get services and information from BMKG. This study only used E-S-Qual without involving E-Recs-Qual indicators in analyzing service quality's impact on customer satisfaction in using the PTSP BMKG application through a questionnaire survey of very limited respondents.

### B. Future Works

Furthermore, research can be continued by analyzing the usability aspect to determine the user experience to improve

PTSP BMKG Application services. In addition, further analysis can also be carried out to find out why this E-S-Qual model does not all dimensions have a positive relationship with customer satisfaction, especially in government services.

## ACKNOWLEDGMENT

Thanks to the PTSP Team and BMKG Indonesia, who have supported and assisted in this research. The Indonesian Ministry of Communication and Informatics Scholarship funds this study. Also, support from the Faculty of Computer Science, University of Indonesia.

## REFERENCES

- [1] R. Rachmawati, D. F. Anjani, A. A. Rohmah, T. Nurwidiani, and H. Almasari, "Electronically-based governance system for public services: implementation in the Special Region of Yogyakarta, Indonesia," *Hum. Geogr.*, vol. 16, no. 1, pp. 71–86, 2022, doi: 10.5719/hgeog.2022.161.5.
- [2] Republik Indonesia, "Peraturan Presiden Republik Indonesia Nomor 61 Tahun 2008 Tentang Badan Meteorologi, Klimatologi, dan Geofisika," Badan Meteorologi, Klimatologi, dan Geofisika, Jakarta, 2008.
- [3] Republik Indonesia, "Peraturan Badan Meteorologi, Klimatologi, dan Geofisika Nomor 5 Tahun 2020 Tentang Organisasi Dan Tata Kerja Badan Meteorologi, Klimatologi, dan Geofisika." Jakarta, 2020.
- [4] A. T. Maulana, "Evaluasi Implementasi Kebijakan Pelayanan Terpadu Satu Pintu (Pfsp) Dalam Upaya Meningkatkan Kepuasan Masyarakat Pada Badan Meteorologi, Klimatologi, dan Geofisika," *J. Signal*, vol. 9, no. 2, p. 296, 2021, doi: 10.33603/signal.v9i2.6283.
- [5] H. B. Abdalla, L. Zhen, and Z. Yuantu, "A New Approach of e-Commerce Web Design for Accessibility based on Game Accessibility in Chinese Market," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 8, pp. 1–8, 2021, doi: 10.14569/IJACSA.2021.0120801.
- [6] H. B. Abdalla, G. Chengwei, and B. Ihnaini, "Improving the Quality of e-Commerce Service by Implementing Combination Models with Step-by-Step, Bottom-Up Approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 9, pp. 17–27, 2021, doi: 10.14569/IJACSA.2021.0120903.
- [7] V. G. de Menezes, G. V. Pedrosa, M. P. P. da Silva, and R. M. d. C. Figueiredo, "Evaluation of Public Services Considering the Expectations of Users—A Systematic Literature Review," *Inf.*, vol. 13, no. 4, pp. 1–14, 2022, doi: 10.3390/info13040162.
- [8] S. Theresia and H. Sen Tan, "Evaluation of service quality and user experience on credit card application using e-SERVQUAL model and usability testing," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 794, no. 1, pp. 1–12, 2021, doi: 10.1088/1755-1315/794/1/012095.
- [9] A. Parasuraman, V. A. Zeithaml, and A. Malhotra, "E-S-QUAL a multiple-item scale for assessing electronic service quality," *J. Serv. Res.*, vol. 7, no. 3, pp. 213–233, 2005, doi: 10.1177/1094670504271156.
- [10] S. D. Kurt and B. Atrek, "The classification and importance of E-S-Qual quality attributes: An evaluation of online shoppers," *Manag. Serv. Qual.*, vol. 22, no. 6, pp. 622–637, 2012, doi: 10.1108/09604521211287589.
- [11] M. Ghosh, "Measuring electronic service quality in India using E-S-QUAL," *Int. J. Qual. Reliab. Manag.*, vol. 35, no. 2, pp. 430–445, 2018, doi: 10.1108/IJQRM-07-2016-0101.
- [12] D. Kang, W. Jang, and Y. Park, "Evaluation of e-commerce websites using fuzzy hierarchical TOPSIS based on E-S-QUAL," *Appl. Soft Comput. J.*, vol. 42, pp. 53–65, 2016, doi: 10.1016/j.asoc.2016.01.017.
- [13] R. R. Ahmed, G. Romeika, R. Kauliene, J. Streimikis, and R. Dapkus, "ES-QUAL model and customer satisfaction in online banking: evidence from multivariate analysis techniques," *Oeconomia Copernicana*, vol. 11, no. 1, pp. 59–93, Mar. 2020, doi: 10.24136/oc.2020.003.
- [14] O. Dastane, M. I. Bin Md Jalal, and K. Selvaraj, "Assessment of Extended E-S-Qual Model in an M-Commerce Setting," vol. 5, no. 12, pp. 923–954, 2018.
- [15] S. L. binti S. A. Kadir, "E-Procurement Service Quality in Malaysia.," *ASEAN Mark. J.*, vol. 8, no. 2, pp. 116–127, 2016, doi: 10.21002/amj.v8i2.9262.

- [16] F. Tjiptono, "Pemasaran Jasa: Prinsip, Penerapan, dan Penelitian (Cetakan 1)." Yogyakarta: Penerbit Andi, 2019.
- [17] W. H. DeLone and E. R. McLean, "The DeLone and McLean model of information systems success: A ten-year update," *J. Manag. Inf. Syst.*, vol. 19, no. 4, pp. 9–30, 2003, doi: 10.1080/07421222.2003.11045748.
- [18] Rissa Hanny and Fahrizal, "Prediction of Online Customer Satisfaction: A Case Study Go-Ride in Jabodetabek," *J. Ekon. Bisnis JAGADITHA*, vol. 8, no. 1, pp. 39–47, 2021, doi: 10.22225/jj.8.1.2758.39-47.
- [19] J. F. Hair, G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, and S. Ray, *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R*. Cham: Springer International Publishing, 2021. doi: 10.1007/978-3-030-80519-7.
- [20] J. F. Hair, J. J. Risher, M. Sarstedt, and C. M. Ringle, "When to use and how to report the results of PLS-SEM," *Eur. Bus. Rev.*, vol. 31, no. 1, pp. 2–24, 2019, doi: 10.1108/EBR-11-2018-0203.
- [21] K. G. Jöreskog, "Simultaneous factor analysis in several populations," *Psychometrika*, vol. 36, no. 4, pp. 409–426, Dec. 1971, doi: 10.1007/BF02291366.
- [22] J. F. Hair, M. C. Howard, and C. Nitzl, "Assessing measurement model quality in PLS-SEM using confirmatory composite analysis," *J. Bus. Res.*, vol. 109, no. December 2019, pp. 101–110, 2020, doi: 10.1016/j.jbusres.2019.11.069.
- [23] A. Sharaf, A. El-Gharbawy, and M. A. Ragheb, "Factors That Influence Entrepreneurial Intention within University Students in Egypt," *OALib*, vol. 05, no. 10, pp. 1–14, 2018, doi: 10.4236/oalib.1104881.
- [24] I. Ghozali, "Aplikasi Analisis Multivariete Dengan Program IBM SPSS 23.," Semarang Badan Penerbit Univ. Diponegoro., vol. Edisi 8, 2016.
- [25] M. Sarstedt, C. M. Ringle, J. Henseler, and J. F. Hair, "On the Emancipation of PLS-SEM: A Commentary on Rigdon (2012)," *Long Range Plann.*, vol. 47, no. 3, pp. 154–160, Jun. 2014, doi: 10.1016/j.lrp.2014.02.007.
- [26] J. Henseler, C. M. Ringle, and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," *J. Acad. Mark. Sci.*, vol. 43, no. 1, pp. 115–135, 2015, doi: 10.1007/s11747-014-0403-8.
- [27] G. Shmueli et al., "Predictive model assessment in PLS-SEM: guidelines for using PLSpredict," *Eur. J. Mark.*, vol. 53, no. 11, pp. 2322–2347, 2019, doi: 10.1108/EJM-02-2019-0189.
- [28] R. T. Cenfetelli, G. Bassellier, and C. Posey, "The analysis of formative measurement in IS research," *ACM SIGMIS Database DATABASE Adv. Inf. Syst.*, vol. 44, no. 4, pp. 66–79, Nov. 2013, doi: 10.1145/2544415.2544420.
- [29] J. F. Hair Jr, G. T. M. Hult, C. M. Ringle, and M. Sarstedt, *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. SAGE Publications, 2021.
- [30] M. Wolfenbarger and M. C. Gilly, "eTailQ: Dimensionalizing, measuring and predicting etail quality," *J. Retail.*, vol. 79, no. 3, pp. 183–198, 2003, doi: 10.1016/S0022-4359(03)00034-4.

# Investigating the Input Validation Vulnerabilities in C Programs

## Input Validation in C

Shouki A. Ebad

Department of Computer Science  
Faculty of Science, Northern Border University  
Arar, Saudi Arabia

**Abstract**—Input validation is a fairly universal programming practice that helps reduce the chances of producing protection-related vulnerabilities in software. In this paper, an experiment is conducted to specifically determine the input validation issues found in programs and the problematic functions that lead to such issues. The experiment evaluated 12 arbitrarily selected open source C projects written by different programmers. The top two most common input validation problems are buffer overflow/XSS and potential memory mismanagement. In addition, the functions that caused the first problem are: (a) strings/text functions (e.g., `strcpy` and `strcmp`), and (b) functions that read from standard input, `STDIN` (e.g., `scanf` and `gets`). In contrast, the functions that caused the second problem are (a) memory allocation/deallocation functions (e.g., `fopen` and `fseek`). Furthermore, the `goto` construct—to a small extent—plays a role. The recommendations are that (a) developers are encouraged to use memory-safe programming languages, otherwise, they should perform different types of checks for the validity of inputs as soon as they are entered, and (b) they should have the required knowledge of secure source code and use tools/suites to manage malformed strings.

**Keywords**—Input validation; buffer overflow; memory mismanagement; safe C functions

### I. INTRODUCTION

Input validation is accepted as good programming practice when writing reliable software. This practice is fairly universal and helps reduce the chances of introducing protection-related vulnerabilities in delivered software [1][2]. This practice can be applied regardless of the programming language (PL) used in development, although the way it is used depends on the specific PL and notations that are used for software development. Every software takes input from its environment and processes it. The specification of such an input makes assumptions about this input that reflects its real-world use. For instance, it may be supposed that an employee ID is always a 10-digit positive integer. However, the software specification does not determine the steps that should be taken in case of wrong input. The user often makes mistakes and sometimes enters the data incorrectly. Regardless of the type of source of unexpected inputs (human, IoT devices, sensors,

the system itself<sup>1</sup>, or other systems), the software can behave in an unanticipated way and provide incorrect values. Such inputs may lead to many security issues. One of these is memory error exploitation, which still ranks among the top three most dangerous software vulnerabilities [1][3].

Buffer overflow and SQL injection are two examples of memory and string-based attacks. The former may be executed using long input strings, and the latter may be performed when a user enters a piece of SQL that is interpreted by a server [4]. The decision that one makes if any validation check fails relies on the software type being developed. For example, while reporting the issue and requesting to re-input the value may be enough in a business application, the input value might have to be estimated according to previous data in a real-time system that should be operated continuously. In contrast, if the source of the input value is a sensor, the most recent valid value could be enough to use. The paper is structured as follows. Section II reviews the existing literature. The research approach is stated in Section III. In Section IV, the results are analyzed and discussed. Section V discusses threats to the study's validity. Finally, Section VI summarizes the article and presents plans for future work.

### II. RELATED WORK

Scholte et al. [4] proposed a technique to prevent the exploitation of cross-site scripting (XSS) and SQL injection vulnerabilities based on the automated data type detection of input parameters. They implemented the technique for PHP and validated it on five web applications with known XSS and SQL injection vulnerabilities. Their technique prevented 83% of SQL injection vulnerabilities and 65% of XSS vulnerabilities while incurring no developer burden. Veen et al. [3] presented memory errors and considered attacks, defenses, and statistics. During a short period in the 70s, they discussed buffer overflows and established CERTs, Bugtraq, and various important methods and countermeasures. A set of related research areas is also explored. Tirronen [6] proposed a technique to eliminate SQL injection attacks by enabling web applications to work with abstract syntax trees while ensuring uniform interpretation of the result. The method involved moving away from processing data as strings to implement a

<sup>1</sup> For configuration, a mobile app can directly read the inputs from itself [5]

non-trivial XSS-protected application with the method in a limited resource. Braz et al. [2] designed an online survey involving 146 participants to understand the extent to which programmers can(not) discover improper input validation vulnerabilities. The participants were assigned to changes with one of two vulnerabilities: (a) SQL injection, and (b) improper validation of specified quantity input. Only 45% of the participants found the vulnerability. There was a lack of knowledge and practices to detect vulnerabilities when an attack scenario is not visible.

Rodríguez et al. [7] analyzed over 50 documents to gather information on the techniques and tools that were used to discover XSS attacks. Their results showed that the trend was increasing in the analysis of content and patterns and decreasing in the use of artificial intelligence to reduce such attacks. Zhao et al. [5] demonstrated an important application of input validation, exposing input-triggered secrets such as backdoors and blacklists of unwanted items. They proposed a tool to find both the execution context of user input validation and the content involved in the validation to expose hidden functionality. The tool was tested with many mobile apps and it was found that they contained more than 12,000 and 4,000 backdoor secrets and blacklist secrets, respectively. Pereira et al. [8] studied a number of buffer overflow vulnerabilities in the Linux kernel, Mozilla, and Xen open-source C/C++ projects to analyze possible methods of improving their detection. The results showed that most of the vulnerable source code units were with defects in checking and dealing with input data types. Static analysis tools lack rules to detect missing or incorrect checking logic vulnerabilities. Moreover, there is no causality between buffer overflow existence and the value of software metrics. Khalaf et al. [9] explored detecting/removing bugs from client-side and server-side code using an input validation mechanism. They supported tests using easy-to-use and accurate models. A program statement that was vulnerable in SQL injection was checked according to static attributes. They also presented a script whitelisting built into the browser's JavaScript engine, where the SQL injection was detected and the XSS attack resolved using that mechanism.

In conclusion, in contrast to the existing literature, this study tries to identify the most common security issues related to input validation and the functions that are the source of such issues. This will help programmers and developers to give more attention to such functions.

### III. EXPERIMENTAL APPROACH

An experiment was performed to understand how well C programs follow input validation practices. The details of the experiment are discussed in the following subsections.

#### A. Research Questions (RQs)

The following RQs are investigated:

RQ1: What are the most common input validation vulnerabilities at source code level?

RQ2: Which functions are the source of the majority of the above input validation vulnerabilities?

#### B. Subjects and Variables

As a representative of software source code, C programs are used; they are collected from the GitHub repository<sup>2</sup>. C is still one of the most common PLs in the market [10]. A controlled experiment is conducted where the subjects of the experiment are 12 open source software (OSS) programs written in C; the subjects are arbitrarily selected. The software developer of the subjects is a factor that could have an impact on the results because individual developers could use the same input validation practices for all of their programs. Therefore, we control this variable by selecting software programs that were designed and developed by different people.

#### C. Data Source and Tools

In order to evaluate the RQs, the data source in our analysis is restricted to the C PL, since (1) it is widely used, and (2) it provides input validation functions or constructs. Additionally, the experiment can be performed without any confounding factors introduced by different PLs. By restricting to just one PL, the results can be placed in context, and we can have more confidence in the conclusions. Table I describes the data source used for this experiment. There is a diversity in domains: education, business, management, tourism, and health. The differences among experimental subjects (i.e., C programs) were not substantial. The well-known metric of non-commented lines of code (LOC) was used to measure the program size. The mean size is 728.2 LOC, indicating the programs are small and medium-sized.

Visual Code Grepper (VCG) was used to collect the input validation violations of C programs. VCG is a source code security tool (available on the web<sup>3</sup>) to handle different programs, including those written in C. The statistical data were analyzed using Microsoft Excel 2016.

## IV. RESULTS AND DISCUSSION

Table II shows the memory-related input validation issues that were discovered by VCG. According to Table II, the top two most common input validation vulnerabilities are potential memory mismanagement and buffer overflow. Before going into the main results in depth, we should first note that the reader is supposed to be familiar with memory and string functions in C. Interested readers can refer to any C textbook for more information.

<sup>2</sup> <https://github.com/>

<sup>3</sup> <https://sourceforge.net/projects/visualcodegrepp/>

TABLE I. DETAILS OF THE EXPERIMENTAL SUBJECTS (C PROGRAMS)

Subject	Project name	Domain	Program name	Program LOC
CP1	Stellarium	Science	Indiserver.c	1528
CP2	Bank-Management-Program	Management	bank management system.c	507
CP3	Departmental-store-management-system	Business	finalProject.c	453
CP4	Library-Management-System	Education	FINAL PROJECT.c	1668
CP5	Calendar	Date and time	Calender&age.c	145
CP6	Contact-Management-System	Management	code.c	204
CP7	Pharmacy management system	Health	Pharmacy Managment System.c	913
CP8	Student-record-system	Education	main.c	943
CP9	Phonebook Application	Management	Phonebook Application.c	288
CP10	Personal Diary Management system	Personal	Personal Diary Managment System.c	618
CP11	Hotel-menu-and-billing-main	Business	main.c	49
CP12	Tux paint	Entertainment	onscreen_keyboard.c	1422
<b>Average</b>				728.2

TABLE II. MAIN ERRORS REPORTED BY VCG

Vulnerability Subject no.	Buffer overflow	Potential Memory Mismanagement	Accepting anonymous internet connection or unverified input data	Can expose residual memory contents	Facilitate format strings bugs	improper control flow	Sum
CP1	48	3	2	5	2	0	60
CP2	42	0	0	0	0	0	42
CP3	28	0	0	0	0	2	30
CP4	25	0	0	0	0	4	29
CP5	4	0	0	0	0	0	4
CP6	18	0	3	0	0	0	21
CP7	29	0	0	0	0	1	30
CP8	107	1	0	0	0	0	108
CP9	2	0	0	0	0	0	2
CP10	32	0	0	0	0	1	33
CP11	6	0	0	0	0	0	6
CP12	23	10	4	2	3	0	42
<b>Sum</b>	364	14	9	7	5	8	407
<b>Mean</b>	30.3	1.2	0.8	0.6	0.4	0.7	33.9

### A. Potential Memory Mismanagement

This problem includes a variety of memory-related errors such as memory leaks, using memory inefficiently, invalid deallocation, double frees, heap corruption<sup>4</sup>, memory

<sup>4</sup> Memory leaks result from memory that is allocated but never freed. Using memory inefficiently happens when a program allocates memory and fails to use it for a long time. It doesn't constitute a memory leak, but can waste a significant amount of memory.

overhead, and file-access violation [3]. The reader is assumed to be familiar with such type of errors; interested readers can consult [3] for more information. Table III shows a sample of lines of code and functions that caused the memory mismanagement problems in C programs. From Table III, the functions and constructs that caused this problem can be divided into three groups:

- Memory allocation/deallocation functions such as memmove, malloc, free, new, and



delete.

- File manipulation functions such as `fopen`, `fprint`, and `fseek`.
- The `Goto` construct.

As a feature in C, the above C programs offered control on memory usage by allowing the optimization of memory allocation for their resources. However, this makes the developers responsible for tracking any memory that their programs dynamically allocate/deallocate. Otherwise, memory-related input validation problems will be introduced as in most cases in Table IV. Some of the above experiment subjects (e.g., CP12) accessed memory via pointers, which produced a memory access error. Use of `goto` should be minimized as much as possible; programmers have been urged to abandon the `goto` statement for more than 50 years on the advice of Dijkstra [11]. Despite this, it is still very much used in C projects [12]. Fig. 1 shows a sample of CP2's source code that uses `goto` in such a case. Use of `goto` came from the fact that C is a non-memory-safe PL in the input validation principle. In particular, it does not have explicit error handling and cleanup constructs like `try/catch` (for exceptions and errors handling) and `finally` (for cleanup activity) in Java; the programmers must therefore resort to using `goto` statements. Most of the current C projects that used `goto` are for these two purposes [12].

In general, tracking memory is not simple—even programs written by skilled programmers contain such problems. The issue originates when an unallocated area is corrupted, and a fatal error often happens in the coming allocation request. Besides the use of `goto` in not handling exceptions and input errors [12], there are three main causes of potential memory mismanagement problems [8].

- Passing a wrong parameter to an allocation function such as `malloc()` and `realloc()`.
- Passing invalid data to a deallocation function such as `free()` and `delete()`.
- Writing before/after the start/end of the allocated space, causing an underrun/overrun error.

### B. Buffer Overflow/XSS

Buffer overflow is a form of memory mismanagement problem. It happens when code goes beyond the portion of data assigned to a buffer. In particular, code with a limited-size buffer accepts unlimited length input. In such a case, the program can crash or malicious code can be executed. In recent years, this issue has grown rapidly with web applications; it is known as XSS attacks, which allow an attacker to insert client-side scripts into web pages that the victim can access; it is also known as internet buffer overflow [9]. As mentioned earlier, writing data to memory beyond a buffer occurs with non-memory-safe PLs like C and C++ that have no bounds checking. Table V shows a sample of lines of code and functions that cause the buffer overflow problem in C programs.

From Table IV, several C functions are known to be unsafe and the source of the vast majority of buffer overflow attacks. They can be divided into two groups:

- Functions to read from STDIN (standard input) such as `scanf()`, `fscanf()`, `sscanf()` and `gets()` where inputs are taken from the keyboard or file.
- Functions to manipulate strings/texts such as `strcpy()`, `strcmp()`, `strlen()`, and `strtok()`.

The advice herein is that C programmers should never use these functions. Fortunately, there are safer alternatives to such unsafe functions. The safer alternatives to `strcpy()` and `strcmp()`, for example, are `strncpy()` and `strncmp()`, respectively. However, the safer functions are not completely safe because `strncpy()` was a cause for buffer overflow in CP1 in Table V. This finding has been confirmed by previous researchers [8]. In particular, the unsafe `strcpy()` takes two arguments—destination and source—and the function copies the source, including the NULL character, to the destination. Contrary to this, the safe `strncpy()` function takes the same two arguments as well as `n`, an unsigned integral type; the function copies the first `n` characters of source to destination i.e., at most, `n` bytes of the source are copied. If there is no NULL character in the first `n` characters of source, the string placed in the destination will not be NULL-terminated. Therefore, there is no guarantee that the destination will be NULL-terminated i.e., a non-terminated string in C is waiting to destroy the program. The question is why such functions were then built in C? The answer comes from C's history; those functions were particularly designed to address specific problems in manipulating strings stored in the manner of original UNIX directory entries, which use a short limited-size array of 14 bytes, and a NULL-terminator was only used when the filename was less than the array.

### C. Timely Recommendations

Although this problem has been known for decades, it is still found in C/C++ software, as has been seen in this study. Any application must not be vulnerable to input validation-based attacks. Therefore there are three timely recommendations herein:

- Use memory-safe PLs. Developers should try not to use non-memory-safe PLs that fail to validate inputs; such a failing can not only lead to buffer overflow attacks (due to long input) but also DoS attacks (due to low memory). In contrast, safe PLs can address these challenges because they check, at runtime, that any access to the memory is within the declared bounds; they remove most buffer overflows at source.
- Perform input validation checks. The first recommendation would not always be a good choice due to the trade-off between performance and security. With memory-safe PLs, there is a necessary performance penalty for this validation, and, for that reason, much code will continue to be written in C. In such a case, the validity of the input should be checked

as soon as it is read. This check includes the specification of the format and structure of the expected inputs, especially considering there are different sources for input, as mentioned in Section 1. Input validation then relies on different checks; used in input checks when the software is implemented, four types of checks [1][5] are shown in Table V.

- Undergo training in writing secure code: The findings have indicated a lack of knowledge and practices to find vulnerabilities. This finding has also been confirmed by a recent study [8] that showed that most buffer overflow vulnerabilities are associated with missing checks (e.g., missing if construct around a statement) or incorrect checking (e.g., the wrong

logical expression used as a branch condition). Regardless of the PLs used in coding, developers should have the required knowledge, training, and practices of secure source code.

- Use appropriate security tools: they should also use static and dynamic analysis security tools that include the use of suites of prebuilt attacks and malformed strings that can quickly discover and eliminate different software vulnerabilities [13]. Examples of such tools that can help developers in this regard include Clang-Tidy, FlawFinder, and Loggly by SolarWinds, which focus on insufficient input validation, XNU memory, and log file analysis and SQL injection, respectively [14].

TABLE III. MAIN ERRORS OF POTENTIAL MEMORY MISMANAGEMENT REPORTED BY VCG TOOL

Subject no.	Problematic statement	Problematic function	Error description
CP1	<code>memmove(ptr, ptr + 1, --len);</code>	<code>memmove</code> <code>malloc</code>	Unrestricted memory copy function. Can facilitate buffer overflow conditions and other memory mismanagement situations.
CP2	<code>goto account_no;</code>	<code>goto</code>	Use of the <code>goto</code> construct. The <code>goto</code> construct can result in unstructured code that is difficult to maintain and can result in failures to initialize or de-allocate memory.
CP3	<code>file2 = fopen("tempfile.txt", "rb");</code>	<code>fopen</code>	Unsafe temporary file allocation. The application appears to build a temporary file with a static, hard-coded name. This causes security issues in the form of a classic race condition (an attacker creates a file with the same name shared between the application's creation and attempted usage) or a symbolic link attack where an attacker creates a symbolic link at the temporary file location.
CP4	<code>rewind(fp);</code>	<code>rewind</code> <code>fopen</code>	The <code>rewind</code> function is considered unsafe and obsolete. Using <code>rewind</code> makes it impossible to determine if the file position indicator was set back to the beginning of the file, potentially resulting in improper control flow. <code>fseek</code> is considered a safer alternative.
CP5, 6, & 11	Buffer overflow (See the next section)		
CP7	<code>fmeds=fopen("Medicines.txt", "r");</code>	<code>fopen</code>	Function used to open a file. Carry out a manual check to ensure that the user cannot modify filename for malicious purposes and that the file is not 'opened' more than once simultaneously.
CP8	<code>new_node = (struct node *)malloc(sizeof(struct node));</code>	<code>malloc</code> <code>fopen</code>	Potential memory mismanagement. Variable name: <code>new_node</code> <code>malloc</code> without free.
CP9	<code>ft=fopen("temp", "wb+");</code>	<code>fopen</code>	Unsafe temporary file allocation. The application appears to build a temporary file with a static, hard-coded name. This causes security issues in the form of a classic race condition (an attacker creates a file with the same name shared between the application's creation and attempted usage) or a symbolic link attack where an attacker creates a symbolic link at the temporary file location.
CP10	<code>rewind(fp);</code>	<code>rRewind</code> <code>fopen</code>	The <code>rewind</code> function is considered unsafe and obsolete. Using <code>rewind</code> makes it impossible to determine if the file position indicator was set back to the beginning of the file, potentially resulting in improper control flow. <code>fseek</code> is considered a safer alternative.
CP12	<code>layout-&gt;keysyms = realloc(layout-&gt;keysyms, sizeof(keysyms) * (i + 1));</code>	<code>realloc</code> <code>malloc</code>	Potential memory leak. On failure, the <code>realloc</code> function returns a NULL pointer but leaves memory allocated. The code should be modified to free the memory if NULL is returned. Dangerous use of <code>realloc</code> : the source and destination buffers are the same. A failure to resize the buffer will set the pointer to NULL, possibly causing unpredictable behavior.

```

add_invalid:
    printf("\n\n\n\t\tEnter 1 to go to the main menu and 0 to exit:");
    scanf("%d",&main_exit);
    system("cls");
    if (main_exit==1)
        menu();
    else if(main_exit==0)
        close();
    else
    {
        printf("\nInvalid!\a");
        goto add_invalid;
    }

```

Fig. 1. Sample of goto procedure in CP2

TABLE IV. MAIN ERRORS OF BUFFER OVERFLOW REPORTED BY VCG TOOL

Subject no.	Problematic LOC	Problematic function	Error description
CP1	strcpy(dp->host, "localhost", MAXSBUF);	strcpy scanf memmove gets strcpy	The function appears in Microsoft's banned function list. Can facilitate buffer overflow conditions. While "safer", the current "n" functions include non-null termination of overflowed buffers and no error returns on overflow.
CP2	scanf("%d/%d/%d",&add.deposit.month,&add.deposit.day,&add.deposit.year);	scanf fscanf	The function directs user defined input to a buffer and so can facilitate buffer overflows.
CP3	strcpy(item.product_code, code);	scanf strcpy	As CP1
CP4	scanf("%d",&i);	scanf gets	As CP2
CP5	scanf("%d",&c);	scanf	
CP6	scanf("%d",&choice);	scanf strcpy	
CP7	gets(c1.name);	gets scanf	
CP8	scanf("%f", &new_node-> university_current_result);	gets scanf	
CP9	scanf("%ld",&p.mble_no);	scanf	
CP10	gets(e.duration);	gets scanf	
CP11	scanf("%d",&a);	scanf	
CP12	#define strtok_r(line, delim, pointer) strtok(line, delim	strtok scanf strcpy	

TABLE V. CHECK TYPES FOR INPUT VALIDATION

#	Check type	Description
1	Range checks	Inputs may be within a particular range. For example, any ratio should be between 0.0 and 1.0; the grade of student should be within the range 0 to 100, the date should be legal (e.g., not February 31st), and so on.
2	Size checks	Inputs are expected to be a given number of characters or upper limit. For instance, an employee ID should be represented with 10 integers, no name with more than 40 characters including family name, no address with more than 100 characters, and so on.
3	Format checks	Inputs may be of specific types; if a number is expected, no alphabetic characters should be allowed. For example, email address should include @ sign, the person's name must be alphabetic with no numbers or punctuation (apart from a hyphen) allowed, and so on.
4	Semantic checks	This check concentrates on the meaning of inputs. As an example, the reading of a household electricity meter should not be so far from that in the corresponding duration in the past year because it is known that the amount of electricity used is expected to be approximately the same.

## V. THREATS TO VALIDITY

Here, we present two threats to the validity of the study's results.

- **Internal validity:** Individual developers would probably carry out the same (insufficiently robust) practices in all programs. This variable was controlled by selecting programs that were written by different developers. In addition, the selection of subjects was arbitrary. The expected threat to internal validity, if there is one, may come from errors in the VCG tool.
- **External validity:** The 12 OSS projects are chosen from different domains to minimize the effect of domain-specific issues. Two factors may affect the interpretation and reduce the generality of the results; studying the input validation practices of developers from 12 OSS projects may not be sufficient, and all projects considered herein are OSS, i.e., not representative of all industrial domains.

## VI. CONCLUSION AND FUTURE WORK

An indicator of secure source code quality is input validation. It is believed that good practices improve program protection, which directly affects recoverability and reliability. In particular, it helps reduce the chances of producing security vulnerabilities in software. This paper conducts an experiment to identify the input validation vulnerabilities in programs and the problematic functions that lead to such issues. The experiment assessed 12 OSS projects written in C, a widely-used PL that provides input validation functions and constructs. The projects have different authors. The results show that buffer overflow (or XSS) and potential memory mismanagement are the top two most common input validation problems. The two types of functions that caused the buffer overflow problem are (a) strings/text functions such as `strcpy` and `strcmp`, and (b) functions that read from standard input, `STDIN`, such as `scanf` and `gets`. In contrast, the functions that caused the memory mismanagement are threefold: (a) memory allocation/deallocation functions such as `memmove` and `malloc`, (b) file manipulation functions such as `fopen` and `fseek`, and (c) the `goto` construct used in handling input errors or exceptions. Two main recommendations are discussed: (a) programmers are encouraged to use memory-safe PLs. Otherwise, they should perform different types of checks for the validity of inputs as soon as they are entered

(four checks are presented in this paper), and (b) in addition they should have the required knowledge of secure source code and should be able to use tools/suites for malformed strings. The results may not be very surprising for skilled C developers, but it is important that there is experimental evidence about the use of a set of C functions and constructs.

There is an open point for further research to examine the problems of using different mechanisms for (a) more than 12 software projects, and (b) real-world systems (not only OSS). However, in the second mechanism, there might be a "data scarcity" research problem due to a lack of sufficient data [14].

## REFERENCES

- [1] I. Sommerville, *Software Engineering*, Pearson, UK, 2015
- [2] L. Braz, E. Fregnan, G. Çalikli, A. Bacchelli, "Why don't developers detect improper input validation? drop table papers", *The IEEE/ACM 43<sup>rd</sup> International Conference on Software Engineering (ICSE)*, Madrid, 22-30 May 2021, DOI: [10.1109/ICSE43902.2021.00054](https://doi.org/10.1109/ICSE43902.2021.00054)
- [3] V. D. Veen, N. Dutt-Sharma, L. Cavallaro, and H. Bos, "Memory errors: the past, the present, and the future", in: Balzarotti, D., Stolfo, S.J., Cova, M. (eds) 'Research in Attacks, Intrusions, and Defenses RAID'. *Lecture Notes in Computer Science*, Volume 7462. Springer, Berlin, Heidelberg, 2012. DOI: [https://doi.org/10.1007/978-3-642-33338-5\\_5](https://doi.org/10.1007/978-3-642-33338-5_5)
- [4] T. Scholte, R. William, B. Davide, and K. Engin, "Preventing input validation vulnerabilities in web applications through automated type analysis", *The Proceeding of the 36th International Conference on Computer Software and Applications*, Izmir, Turkey, pp. 233-243, 16-20 July 2012, DOI: [10.1109/COMPSAC.2012.34](https://doi.org/10.1109/COMPSAC.2012.34)
- [5] Q. Zhao, C. Zuo, B. Dolan-Gavitt, G. Pellegrino, Z. Lin, "Automatic uncovering of hidden behaviors from input validation in mobile apps", *The IEEE Symposium on Security and Privacy (SP)*, San Francisco, CA, USA, 18-21 May 2020, DOI: [10.1109/SP40000.2020.00072](https://doi.org/10.1109/SP40000.2020.00072)
- [6] V. Tirronen, "Stopping injection attacks with code and structured data", in M. Lehto, & P. Neittaanmäki (Eds.), *Cyber security: power and technology*, 2018, pp. 219-231. Springer. *Intelligent Systems, Control and Automation: Science and Engineering*, 93. DOI: [https://doi.org/10.1007/978-3-319-75307-2\\_13](https://doi.org/10.1007/978-3-319-75307-2_13)
- [7] G. E. Rodríguez, J. Torres, P. Flores, D. E. Benavides, "Cross-site scripting (XSS) attacks and mitigation: a survey", *Computer Networks*, Volume 166, 15 January 2020, 106960. DOI: <https://doi.org/10.1016/j.comnet.2019.106960>
- [8] J. D. Pereira, N. Ivaki, and M. Viera, "Characterizing buffer overflow vulnerabilities in large C/C++ projects", *IEEE Access*, vol. 9, 2021, DOI: [10.1109/ACCESS.2021.3120349](https://doi.org/10.1109/ACCESS.2021.3120349)
- [9] O. I. Khalaf, M. Sokiyna, Y. Alotaibi, A. Alsufyani, and S. Alghamdi, "Web attack detection using the input validation method: DPDA theory", *Computers, Materials & Continua*, vol. 68, no. 3, 2012, DOI: [10.32604/cmc.2021.016099](https://doi.org/10.32604/cmc.2021.016099)
- [10] S. A. Ebad, A. A. Darem, and J. H. Abawajy, "Measuring software obfuscation quality—a systematic literature review", *IEEE Access*, vol. 9, 2021, 99024–99038.

- [11] E. W. Dijkstra, "Goto statement considered harmful", Communications of ACM: Letters to the editor, 1968, vol. 11, no. 3, pp. 147–148.
- [12] M. Nagappan, R. Robbes, Y. Kamei, É. Tanter, S. McIntosh, A. Mockus, A.E. Hassan, "An empirical study of goto in C code from GitHub repositories", The Proceedings of the ESEC/FSE'15: Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering, Bergamo, Italy, Aug. 30- Sep. 4, 2015, DOI: <https://doi.org/10.1145/2786805.2786834>
- [13] A. Rashid, H. Chivers, G. Danezis, E. Lupu, A. Martin, "The Cyber Security Body of Knowledge", CyBOK Version 1.0. 2019. [Online]. Available: <https://www.cybok.org/media/downloads/CyBOK-version-1.0.pdf>, accessed 20 Jan 2023.
- [14] M. Taeb, and H. Chi, "A personalized learning framework for software vulnerability detection and education", International Symposium on Computer Science and Intelligent Controls (ISCSIC), 12-14 November 2021, Rome, Italy. DOI: <https://doi.org/10.1109/ISCSIC54682.2021.0003>
- [15] S. A. Ebad, "An exploratory study of ICT projects failure in emerging markets", Journal of Global Information Technology Management, vol. 21, no. 2, 2018, pp. 139-160. DOI: <https://doi.org/10.1080/1097198X.2018.1462071>

# An Optimized Method for Polar Code Construction

Issame El Kaime<sup>1</sup>, Reda Benkhouya<sup>2</sup>, Abdessalam Ait Madi<sup>3</sup>, Hassane Erguig<sup>4</sup>

Advanced Systems Engineering Laboratory-National School of Applied Sciences, Ibn Tofail University Kenitra, Morocco<sup>1</sup>  
MISC Laboratory-Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco<sup>2</sup>

Advanced Systems Engineering Laboratory-National School of Applied Sciences, Ibn Tofail University Kenitra, Morocco<sup>3</sup>  
Materials Physics and Subatoms Laboratory-Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco<sup>4</sup>

**Abstract**—Polar codes are traditionally constructed by calculating the reliability of channels, then sorting them by intensive calculations to select the most reliable channels. However, these operations can be complicated especially when, the polar code length,  $N$  becomes great. This paper proposes a new low-complexity procedure for polar codes construction over binary erasure and additive white Gaussian noise (AWGN) channels. Using the proposed algorithm, the code construction complexity is reduced from  $O(N \log N)$  to  $O(N)$ , where  $N=2^n$  ( $n \geq 1$ ). The proposed approach involves storing the classification of channels by reliabilities in a vector of length  $L$ , and then deriving the classification of  $M$  channels for every  $M$  where  $M \leq L$ . The proposed method is consistent with Bhattacharya parameter based Construction and Density Evolution with Gaussian Approximation (DEGA) based construction. In this paper, the Successive Cancellation Decoding algorithm (SCDA) is used. Thanks to its low complexity and its high error-correction capability.

**Keywords**—Polar codes; SNR; successive cancellation decoding; error correction; low-complexity; code construction; additive white Gaussian noise; Bhattacharya parameter; density evolution with Gaussian approximation

## I. INTRODUCTION

It is usual for communication links to suffer from errors due to random noise, interference and malfunctioning devices, etc. To correct errors in channel coded data streams, a set of algorithmic operations is applied to the original data stream at the transmitter. A second set of algorithmic operations is applied to the received data stream at the receiver. Encoding and decoding operations at the transmitter and receiver are collectively called channel coding operations in channel coding terminology. Research in channel coding is focused on developing high performance channel codes that mitigate the effects of errors in communication links. A real challenge here is to accomplish this with sufficient simplicity to allow practical implementation in silicon technology. Everything depends on the complexity of a code, including its power consumption, memory requirements, computation power requirements, and latency, which determine whether or not a code is appropriate for any given scenario. Channel coding is somewhat revolutionized by polar codes.

The polar codes proposed by Arikan in 2008 can achieve that capacity of any binary discrete memoryless channel [1], researchers from all over the world have been interested in polar codes ever since their introduction in [1], polar codes can be used in lot of applications like cryptography [2], speech communication [3], data storage [4]. Polar codes are also used

to develop new block [5]. Polar codes are composed of three main stages; namely the construction, the coding and finally the decoding. The construction of polar codes is a crucial step as they affect the performance of polar codes. In polar codes construction, synthetic channels are evaluated by reliability. Good channels are selected for information transmission and the bad channels are frozen. Where polar codes are capable to achieving channel capacity for any binary-input discrete memoryless channel [1].

Polar codes construction step encounters the following problem: given  $N$  code lengths and  $K$  information bit length, what is the best way to select  $K$  channels out of all the available ones, knowing that the remaining  $N-K$  bit channels are frozen and provided to the transmitter and receiver. An indication of the quality of a virtual bit channel  $W_N^i(\cdot)$  can be determined by using a variety of metrics. Polar codes take effect when the length of the code is very large, which implies a large number of computes to know the capacity of each channel, As a result, building polar codes is extremely difficult and requires a lot of resources. This paper considers the construction of polar codes over symmetric binary discrete memory-less channels by using a new method to reduce the complexity of the construction. Many methods have been developed previously in literature to build polar codes, Monte-Carlo simulations are proposed in [1] with a high complexity of  $O(TN \log N)$  where  $T$  indicates the number of iterations of Monte-Carlo simulations. In [6] and [7], polar codes construction is based on density evolution, where convolutions of functions are performed and numerical calculation precision is limited by the complexity of the process. In [8] bit-channel approximations are proposed with a Complexity under controlled conditions of  $O(N \cdot \mu^2 \log \mu)$  ( $\mu$  a user-defined parameter that limits the number of output alphabets at each step of the approximation process). Another type of algorithm can construct polar codes using Gaussian approximation (GA) of additive white Gaussian noise (AWGN) channels [9]–[11], this approximation function [11] inherently limits the GA method, with some restrictions on the length of blocks [11]. Bhattacharya parameters are used to construct polar codes in [1], other constructions methods with variable performance and complexity are located in [12]–[14].

This paper describes an efficient method of constructing polar codes to reduce their computational complexity, if the method suggested in this paper is compared to other ones in the literature, the method presented here is characterised by a reduced complexity.

In the following sections of this paper, polar codes are described, including their background, with a focus on the concept of channel polarization. Section III presents the traditional method of constructing polar codes, while Section IV provides numerical results that demonstrate the effectiveness of the proposed method compared to traditional methods and state-of-the-art techniques. Finally, the conclusion summarizes the key findings of the study and discusses potential future research directions.

## II. POLAR CODES

In general, all channel coding technologies work in quite similar ways, even if the excellent performance of turbo codes and LDPC (Low Density Parity Check) codes in practice, none of the last codes can be demonstrated to attain the capacity of channels exempt the binary erasure channel (BEC). Polar codes are members of the block code family since they operate on blocks of symbols/bits. To construct polar code, two key operations are required: channel combining and channel splitting. The channel combining process, carefully selected combinations of bits are mapped to specific channels.

Let consider N bits to be sent over discrete binary channel without memory (B-DMC) W. Each transmission represents a use of W, which means that each bit is passed through a copy of W as shown in Fig. 1.

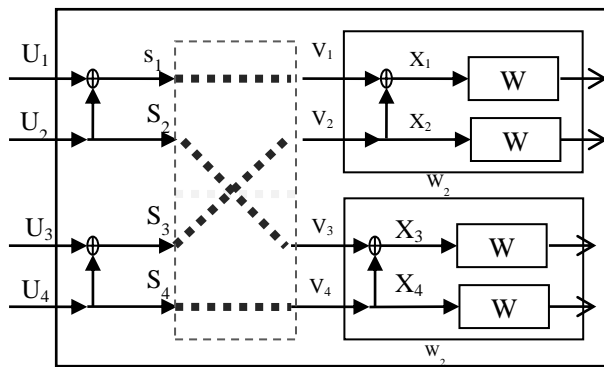


Fig. 1. Combination of two copies of W to form W2.

The map in Fig. 1 between the input  $u_2^1$  and the vector  $X_2^1$  can be represented by the equation (1):

$$X_2^1 = u_2^1 * G_2 \tag{1}$$

Where  $G_2$  is the basic matrix:

$$G_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \tag{2}$$

The same operation is repeated with  $W_2$  as the basic element to produce  $W_4$  like presented in Fig. 2. Generally At this step, the combined channels  $W_N$  is the virtual channel that map the input data  $u_N^1$  to the output  $y_N^1$ .  $W_N^i$  are divided into a set of N bit input channels  $W_N^i()$ .

These virtual channels  $W_N^i()$  are characterized by the transition probabilities provided by equation (3).

$$W_N^{(i)}(y_1^N, u_1^{i-1} | u_i) \triangleq \sum_{u_{i+1}^N \in X^{N-i}} \frac{1}{2^{N-i}} W_N(y_1^N | u_1^N) \tag{3}$$

$$(W_N^{(i)}) : X \rightarrow Y^N \times X^{i-1}, 1 \leq i \leq N \tag{4}$$

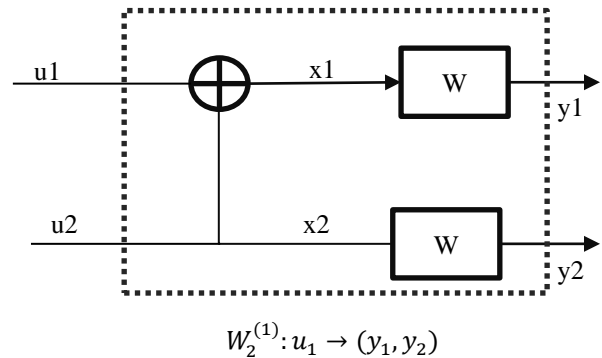


Fig. 2. Combination of two copies of W2 to form W4.

## III. THE NATIVE CONSTRUCTION

As mentioned earlier, building polar codes is like finding all the best positions of reliable bits positions. The selection of the information bits is one of the most important steps in polar coding. Polar codes are originally constructed using simple bounds on Bhattacharyya parameter bit channels [1], The Bhattacharyya parameter is linked to the Bhattacharyya distance, which measures the similarity of the probability distributions of two symbols [15]. It has been widely used to produce good polar codes because of its simplicity. By creating a system allowing one to access each bit-channel individually, one can send data only through those for which the Bhattacharyya parameter is close to 0. Bhattacharyya's parameter  $Z(\cdot)$  defined in (5) provides an upper bound on the error probability of transmission over W with maximum likelihood (ML) decisions when the channel is used only once.

$$Z(W_N^{(i)}) = \sum_{y_1^N \in Y^N} |X_{i+r}^N \cdot y_1^N| \cdot \sqrt{W_N^{(i)}(y_1^N, 0_{i-1} | 0) \cdot W_N^{(i)}(y_1^N, 0_{i-1} | 1)} \tag{5}$$

Accordingly, channels with  $Z(W_N^i) < \epsilon$  are almost noiseless, whereas channels with  $Z(W_N^i) > 1 - \epsilon$  are almost pure-noise channels where  $0 < \epsilon < 1$  [15]. However, it has been cited that the parameters' updates succeeded with equality just for the Binary Erasure Channel.

### A. Density Evolution with Gaussian Approximation (DEGA) Construction

Over AWGN channel the Density Evolution with Gaussian Approximation (DEGA) is the famous construction for polar codes. From the channel stage towards the decision stage, DEGA attempts to evolve the densities of the LLRs via the decoder. Regardless, precise if the begin is with Gaussian densities at the channel stage, the consequent densities at the following stages are not Gaussian anymore. The DEGA decoder relaxes this by supposing they are approximately Gaussian, and consequently it only tracks their mean and variance throughout the approach. The relationship between

mean and variance should survive at all stages [10]. The variance and the mean is defined in (6) where the LLR is defined in (7).

$$m = E\{L\} = \frac{2}{\sigma^2}, \quad \text{var}\{L\} = \frac{4}{\sigma^2} = 2m \quad (6)$$

The next update is accomplished by DEGA[9].

$$m_i^{(s)} = \begin{cases} \phi^{-1}\left(1 - \left(1 - \phi(m_i^{(s-1)})\right)^2\right), & \text{if node } f, \\ 2m_i^{(s-1)}, & \text{if node } g, \end{cases} \quad (7)$$

The function  $\phi$  is given by (8).

$$\phi(x) = \begin{cases} \exp(0.4527x^{0.86} + 0.0218), & 0 < x < 10 \\ \sqrt{\frac{\pi}{x}} \exp\left(-\frac{x}{4}\right) \left(1 - \frac{10}{7x}\right), & x \geq 10 \end{cases} \quad (8)$$

#### IV. NUMERICAL RESULTS AND DISCUSSION

This section analyses the classifications of polar codes virtual channels by reliability for different code lengths over AWGN and BEC, the information length is not tested because the construction of polar codes depends on the code length.

##### A. Classification of Polarizing Channels over BEC

This part starts by the comparison by reability of 8 BEC polarizing channels with erasure probability 0.1,0.2,0.3,0.4,0.5,0.6,0.7 and 0.8, the results are shown in Fig. 3.

The simulated results that are calculated using the relation (17) of [1] indicates that virtual channels are classified in the same way independently of the erasure probability. Fig. 3 shows that for all erasure probabilities, virtual channels are classified as below 8->7->6->4->5->3->2->1 (from the most reliable to the least accessible).

Fig. 4 shows the reliability of 16 virtual channels for the following erasure probabilities [0,1,...,0,9]. Fig. 4 shows that the classification of 16 virtual channels is the same for all erasure probabilities, the classification by consequence is

16-> 15->14-> 12-> 8-> 13-> 11-> 10-> 7-> 6-> 4-> 9-> 5->3->2->1.

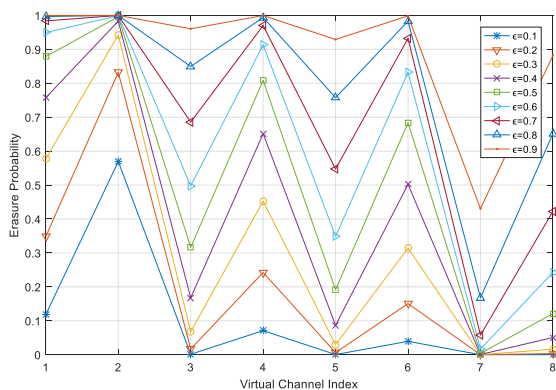


Fig. 3. The Bhattacharyya for 8 BEC polarizing channels.

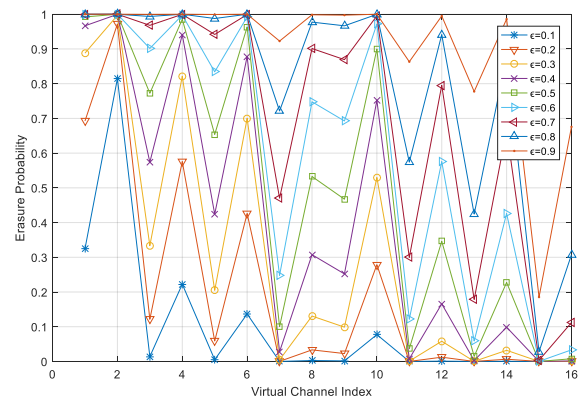


Fig. 4. Comparison of 16 BEC polarizing channels in terms of reliability with erasure probability [0.1, 0.2, ..., 0.9].

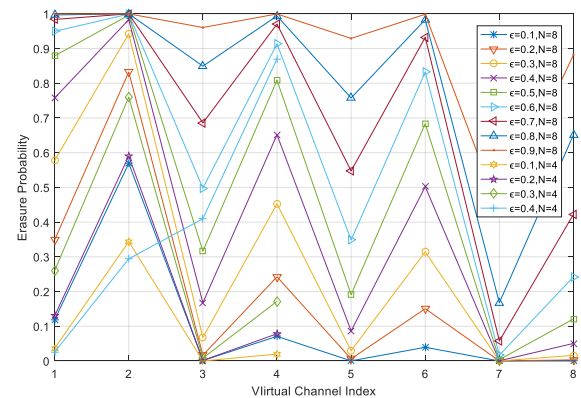


Fig. 5. Classification of channel by reliability for N=8 and N=4.

Fig. 5 compares the classification of virtual channels for different code lengths N=8 and N=4.

The results shown in Fig. 5 proof that the classification of virtual channels with index less than 4 is the same in both codes.

Table I provides the classification of virtual channels by reliability for several code lengths (N=1,2,4,8,16,32) where the erasure probability  $\epsilon=0,25$ . For ease of analysis, each channel index is characterized by the same color in each code length. The table shows that the classification is the same when the length of the code changes.

##### B. Classification of Polarizing Channels under AWGN

This section analyses channel classifications by reliability for different code lengths over AWGN, the Bhattacharyya parameters are used to evaluate the reability for each virtual channel, note that the same tests can be applied with Gaussian density approximation method.

In the beginning, reliability of eight virtual channels are compared under different values of signal to noise ratio (SNR). Fig. 6 shows the reliability of eight polarizing channels using SNR -2, -1, 0, 1, 2 and 3. It also worth nothing that the classification of the eight polarizing channels by reliability remains the same when SNR changes.





The Table II generalizes the simulation of Fig. 9 by presenting the classification of virtual channels by reliability for several code lengths ( $N=2,4,8,16$ ) where the  $SNR=2$ . For ease of analysis, each channel index is characterized by the same color in each code length. The table shows that the classification is the same when the length of the code changes.

Mathematically the proposed procedure to construct polar codes with low complexity will be described below in algorithm 1. This algorithm uses linear search where every element in a reliability vector is checked, starting at 0 and going through each element until the desired element is found. Otherwise, the search continues until the end of the list.

---

**Algorithm 1:** Find out all good channels index and bad ones

---

**Input :**  
Channel index vector classified by reliability  $I$  with length  $N_i$   
polar code length  $N$  where  $N \leq N_i$

**output:**  
 $S$ : virtual channel classified by reliability for polar code with length  $N_r$

```
1: for k=0 to N do
2:   for J=0 to  $N_i$  do
3:     if the value of  $N_i[J]$  less or equal N
4:        $S[k]= N_i[J]$ 
5:     endif
6:   endfor
7: endfor
8: Return S;
```

---

The latest algorithm uses linear search where each item in a list is checked, starting at 0 and going through each item until the desired item is found. Otherwise, the search continues until the end of the list. There is no better search algorithm. The complexity of the newly developed build algorithm is the same as the complexity of the search algorithm, which is  $O(N)$ .

To better understand the algorithm designed, let's take an example of building polar codes, assumes that Gaussian density was adopted firstly to construct polar codes. The new proposed algorithm for constructing polar codes can be applied here by storing in a particular vector ( $R$ ), the classification of virtual channels by reliability from the least reliable to the most reliable. Next the reliability calculation for each virtual channel will not be done in the next polar code construction, there is a need to select the most reliable channels of the  $R$  vector with an index less than the length of the new polar codes.

## V. CONCLUSION

The previous approaches of polar codes construction require one to one compute of the reliability for all synthetic channels and use only those that are sufficiently reliable [2]. For the gain in complexity and resources, it is worthwhile to

perform an optimized construction polar codes algorithm. the proposed method decreases the computation complexity in the construction of Polar Codes. Clearly, the complexity is reduced to  $O(N)$ . Note that when using this optimized construction approach, extra memory is allocated to store the vector of reliability.

The memory resources play a crucial part in the outcome of this study, these limitations do have a significant impact on the primary finding, future research could seek to reduce memory's impact by choosing the best value of the vector reliability length for every application.

## REFERENCES

- [1] E. Arıkan, « Channel Polarization: A Method for Constructing Capacity-Achieving Codes ».
- [2] J. Liu, Y. Wang, Z. Yi, et Z. Lin, « polarRLCE: A New Code-Based Cryptosystem Using Polar Codes », Security and Communication Networks, vol. 2019, p. e3086975, 2019, doi: 10.1155/2019/3086975.
- [3] S. Zhao, P. Shi, et B. Wang, « Polar codes and its application in speech communication », in 2011 International Conference on Wireless Communications and Signal Processing (WCSP), 2011, p. 1-4. doi: 10.1109/WCSP.2011.6096731.
- [4] K. N. Tunuguntula, « Polar codes for data storage and communication network applications », UC San Diego, 2022.
- [5] D. Khebbou, R. Benkhouya, I. Chana, et H. Ben-azza, « Finding Good Binary Linear Block Codes based on Hadamard Matrix and Existing Popular Codes », IJACSA, vol. 12, no 11, 2021, doi: 10.14569/IJACSA.2021.0121150.
- [6] R. Mori et T. Tanaka, « Performance and construction of polar codes on symmetric binary-input memoryless channels », in 2009 IEEE International Symposium on Information Theory, Seoul, South Korea, 2009, p. 1496-1500. doi: 10.1109/ISIT.2009.5205857.
- [7] R. Mori et T. Tanaka, « Performance of Polar Codes with the Construction using Density Evolution », IEEE Commun. Lett., vol. 13, no 7, p. 519-521, 2009, doi: 10.1109/LCOMM.2009.090428.
- [8] Ido Tal and Alexander Vardy, How to Construct Polar Codes.
- [9] P. Trifonov, « Efficient design and decoding of polar codes », IEEE Transactions on Communications, vol. 60, no 11, p. 3221-3227, 2012.
- [10] D. Wu, Y. Li, et Y. Sun, « Construction and block error rate analysis of polar codes over AWGN channel based on Gaussian approximation », IEEE Communications Letters, vol. 18, no 7, p. 1099-1102, 2014.
- [11] J. Dai, K. Niu, Z. Si, et J. Lin, « Evaluation and optimization of Gaussian approximation for polar codes », in Proceedings of the American Society for Information Science and Technology, 2016, vol. 51, no 1, p. 1-2.
- [12] R. Pedarsani, S. H. Hassani, I. Tal, et E. Telatar, « On the construction of polar codes », in 2011 IEEE International Symposium on Information Theory Proceedings, 2011, p. 11-15.
- [13] D. Kern, S. Vorkoper, et V. Kuhn, « A new code construction for polar codes using min-sum density », in 2014 8th International Symposium on Turbo Codes and Iterative Information Processing (ISTC), Bremen, Germany, 2014, p. 228-232. doi: 10.1109/ISTC.2014.6955119.
- [14] H. Li et J. Yuan, « A practical construction method for polar codes in AWGN channels », in IEEE 2013 Tencon-Spring, 2013, p. 223-226.
- [15] R. Benkhouya, I. Chana, et Y. Hadi, « Study of the operational SNR while constructing polar codes », IJECE, vol. 10, no 3, p. 3200, 2020, doi: 10.11591/ijece.v10i3.pp3200-3207.

# Current Multi-factor of Authentication: Approaches, Requirements, Attacks and Challenges

Ali Hameed Yassir Mohammed<sup>1</sup>, Rudzidatul Akmam Dziauddin<sup>2</sup>, Liza Abdul Latiff<sup>3</sup>  
Razak Faculty of Technology and Informatics-University Technology Malaysia, Kuala Lumpur, Malaysia<sup>1,2,3</sup>  
College of Computer Science and Information Technology-University of Sumer, ThiQar, Iraq<sup>1,2</sup>

**Abstract**—Now-a-days, with the rapid and broad emergence of local or remote access to services on the internet. Authentication represents an important security control requirement and the MFA is recommended to mitigate the weaknesses in the SFA. MFA techniques can be classified into two main approaches: based biometric and non-biometric approaches. However, there is a problem to maintain the tradeoff between security and accuracy. The studies that have been reviewed on both authentication mechanisms are found contradictory in the direction of others. In the direction of authentication-based biometrics the researchers tended to increase the recognition accuracy, while in the other direction, the researchers proposed to combine many authentication factors to increase the security layers. The main contribution of this survey is to review and spotlight on the current state of the arts in both authentication mechanisms to achieve a secure user identity. This paper provides a review of authentication protocols and security requirements. In addition to a detailed review with a comparison of secure one-time passcode generation and distribution. Furthermore, a comprehensive review of cancelable biometrics techniques, attacks, and requirements. Finally, providing a summary of key challenges and future research directions.

**Keywords**—MFA; authentication; OTP; cancelable; biometrics; security; identity

## I. INTRODUCTION

Lately, organizations worldwide have made a quantum leap in terms of online application as services expand to the customers' satisfaction [1]. This led to the manifestation of widespread modernized applications as example, users can easily perform various operations such as withdrawal, payment, and transfer over internet banking websites [2]. With the increment of using smartphone devices, authentication protocols can be classified into two approaches non-biometrics-based approach and based biometrics approach [3].

Online identity access management (IAM) is one of the most important services that require customer authentication during daily transactions of funds in a secure way [4]. Some current surveys emphasized various types of customer authentication protocols used for secured data transfer including Single Factor Authentication (SFA) and Multi-Factor Authentication (MFA). For instance, [5] elaborated in detail the significance of various factors used for authentication and compared these factors based on different parameters such as universality, uniqueness, collectability, performance, and usability. The evolution mechanism of the authentication systems is explained in [6] where MFA was implemented for

the user and Vehicle-to-Everything (V2X) interactions. A framework was proposed to identify the missing factors and further authenticate the users without supplying any sensitive biometric information. Consequently, it enabled an elastic in-car verification of the occupant using efficient integrated sensors.

In [1] authors analyzed the situation of 30 banks regarding the MFA execution in online banking, wherein the main aim was to determine the impact of the MFA protocols on the regulations, practices, and system security against attacks in the banking sector. It was acknowledged that although it is difficult to implement the MFA systems, wide adoption of the validators taking advantage of the inheritance factors can improve both the security and efficiency of the MFA systems. Broader adoption of authenticators that take the advantage of inheritance factors can improve the security and intricacy of the MFA systems. In the context of mobile devices, a comprehensive survey was conducted on the behavioral patterns of biometrics and constant verification methods [7]. In addition, the behavioral biometrics and verification of the mobile device; various methods of behavioral biometrics and feature extraction were analyzed with a focus on machine learning (ML) models' performance. The limitations of the ML models were discussed with respect to their usage due to the security, and usability as well as privacy concerns.

Also, [8] reviewed the current trends of various MFA protocols and analyzed the gaps in the current literature for future studies related to the perception of users' risks. Different identifiable trends in the MFA studies were found, indicating the need for new validation techniques. However, it lacked risk perception analysis. This work disclosed the presence of cultural and demographic biases in the user study designs. In general, a recruitment bias for the users was achieved in the context of an academic background.

While [9] argued that an increase in the attacks against the MFA mechanism is mainly related to the nature of the authentication protocols, factors, and their importance for the security of the advent of mobile money. The study conducted a literature review of the attacks and countermeasures in the MFA for mobile money. The authors recommended for a future authentication system use a secure multifactor authentication scheme Personal Identification Number (PIN), One Time Passcode (OTP), and biometrics features. Moreover, protect data during distribution and storage using end-to-end encryption methods. Furthermore, it was stated that despite the coverage of several attacks against mobile money other attacks

must be considered. Some of best practices are recommended by [10] to overcome the drawbacks of several mitigation strategies. Meanwhile, the implementation of many countermeasures like new families of hash functions was recommended for security. The study [11] presented a survey of the Cancelable Biometric (CB) methods of various types based on cryptography, filtering, transformation, multi-models, and hybrid methods. It also addressed requirements and performance measures as well as attacks and challenges.

Most of the related surveys are in the field of non-biometric factors and concern the issues related to the improvement of the authentication by improvement of MFA-based OTP generation and distribution. In the second direction, the studies tackle the authentication-based current trend of cancelable biometrics methods. Secure digital identity management is the main aspect of any authentication system that cannot be ignored [12].

Thus, this study refers to the most common performance metrics, requirements, and attacks for both approaches of the related studies [1], [4], [5], [6], [7], [8], [9], [10], and [11].

Although multifactor authentication is a growing and prospective research scope but lacks a comprehensive survey on this field is not available except for fewer research. This survey has made the following contributions:

- The paper has presented a comprehensive survey of both multi-factor of authentication approaches and the state of arts; MFA-based biometrics and non-based biometrics.
- Comprehensive review of various security attacks and performance measures used in CB.
- Comprehensive review of various security attacks and performance measures used in non-based biometrics approaches.

Section II provides a review of identity access management and discusses the authentication protocols, authenticators' classification, Authentication security requirements, and attacks with countermeasures. Section III presents MFA features, Biometrics types, requirements, attacks, and performance metrics of conventional and current trends of cancelable biometrics, and Section IV provides a review of the current state of the arts of cancelable biometrics approaches. Section V describes the authentication of non-based biometrics and provides a security analysis of the current OTP generation and distribution methods. Section VI emphasizes the key challenges. Section VII concludes the paper and presents the future directories.

## II. IDENTITY ACCESS MANAGEMENT PROTOCOLS

An Identity Access Management Protocol (IAM), is a controlled access that manages the identity provider (Idp), controls the client authentication (signed in), and authorizes (permissions) the use of the targeted resources. IAM organizes the resource's availability, and accessibility and preserves data privacy [4].

IAM protocols permit the digital verification of a client based on several factors. The recent advancement in innovative

communication, security authentication, computerized e-payment, and smartphone devices faced increasing authentication challenges due to diverse security threats from attackers or phishers [13].

A conventional SFA approach can keep approved access, login, or get to the secret content by utilizing a single factor-like username and password. Now, this approach was hardened by utilizing at least two factors in the combination process [14]. According to National Institute of Standards and Technology (NIST), the digital identity of a user is defined as a series of functions of authentication, authorization, and accounting of a client in a specific context in a unique way (for example a payment service) [4].

The authentication of the digital identity of a user is achieved in practice via the so-called authentication protocol to get authorization to get access granted to the resources (data, computer program, call object, or procedure). The access server is explained in [1] where requests additional user information for example in the earlier application layer protocol Remote Authentication Dial-In User Service protocol (RADIUS), OpenID, Kerberos, OAuth delegation framework.....etc. Modern identity providers (Idp) rely on Role-Based Access Control (RBAC) to access resources.

Access control utilizes authentication to verify the user identity (uid). In addition, the accounting process is the job to manage and keep the records of the user or any other related objects and works to provide an assertion of the authorized object to the Service Provider (SP) [15].

Furthermore, there are many protocols used to manage the assertion or to transfer the attributes pairs names and values like Security Assertion Markup Language (SAML) protocol. SAML protocol transfers the authentication, attribute, and authorization decision statements from the Idp to SP to perform an action on the requested resource among the relying parties as in Single Sign on (SSO) protocol [16].

SSO is supported by Azure Active Directory (Azure AD) SAML authentication sends requests AuthnRequest (authentication request) and receives the responses from the Idp (Azure AD), Fig. 1 describes the SSO processes sequences to use HyperText Transfer Protocol (HTTP) to post and bind the responding to the SP [16].

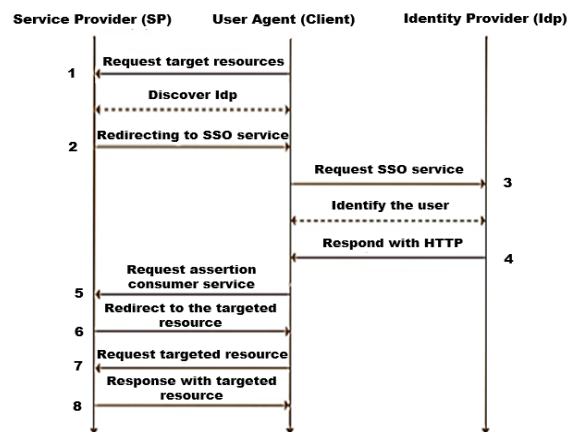


Fig. 1. Single sign on workflow.

### A. Authenticators Classification

Authenticators can be classified into different categories such as memorized and look-up secrets, out-of-band devices, token devices, and software as shown in Fig. 2. The salient attributes of these authenticators are briefly described hereunder [1].

1) *Memorized secret*: When a client shares memorable pieces of information with the server he/she has to recall including passwords, PINs, pass-phrases, and secret questions. The old methods keep the memorized secrets through encryption or salting. However, according to the recommendations of NIST, the memorized secrets must be stored in a form that can withstand both online/offline attacks. In addition, it must be hashed via the approved hash function [17].

2) *Look-up secret*: It can be electronic records of a secret set shared between the client and server to attest to a possession factor. Then to complete the transaction, the client has to provide the secret associated with definite positions in the records [18]. For example, the user or applicant may be asked by the verifier to give certain subsets of the alphanumeric characters or strings printed on a card in a tabular format [19].

3) *Authenticator device*: A small hardware device is used to generate an authenticator output. This authenticator platform may be constructed into a specific user's device and employed on the connected devices whereas the roaming verifier links to a device platform via transport protocols [20].

4) *Software authenticator*: It is sometimes called a software token, wherein the programs are executed to generate the authenticator's output. The software authenticator act as the coordinate of verifier devices for both SFA and MFA software authenticators. The authenticator based on the software may be implemented on laptops, tablet computers, or smartphones. For instance, a mobile application on the user's smartphone can be considered a kind of phone-based authenticator [21]. For security purposes and to prevent unauthorized access to the private or secret data domain, an authenticator based on the software might employ the Trusted Platform Module (TPM) on the client device or trusted execution environments of the processor. Authenticator devices or hardware are classified into SFA and MFA types as explained below [22]:

a) MFA is known as a Time based OTP key (TOTP) token. When the TOTP devices or software share the user's OTP periodically, OTP will be popped up to the user or reach his email or SMS box. Then, the user should manually enter the OTP into the input text box for the successful execution of an MFA protocol completely [22].

b) Using MFA, various devices or software generate tokens in the form of alpha-numeric string which needs some PIN, biometric or secret data for activation, thereby confirming both ownership [22].

5) *Out-of-band device*: These types of authenticators are shared exclusively over the specific secondary channels (for

example an SF device) that approve the possession factor initiated by a secondary mobile phone network channel. Usually, this kind of authenticator depends on a SIM card [4]. The MFA methods rely on mobile phones app where some depend on notification or OTP authentication (event and time-based). Upon receiving the OTP generated by the server and dynamical sharing the user can go for the online transaction [17]. The SMS notification is subjected to security concerns as analyzed afterward. The main flaw in the authentication process is that the user must carry many hard tokens for a different account like Universal 2<sup>nd</sup> Factor of authentication (U2F), Near Field Communication (NFC) devices, or Secure Digital (SD) cards. Another threat to the authentication of the user's account is related to the loss or theft of these devices. In addition, many organizations prohibit carrying the electronic devices to or outside the workplace due to security policies. Due to this reason, most of the mobile phones do not have flash memory ports. One cannot upgrade or update the hard tokens because they require a new hard token for each account up-gradation or reconfiguration [2]. High costs are involved in purchasing or exchanging tokens of this type.

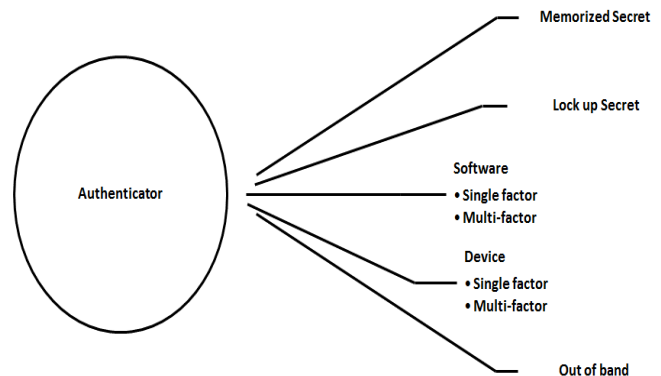


Fig. 2. Authenticator's classification.

The MFA scheme for mobile and smartphones gives an option for the dedicated physical device. For the authentication process, one can use the security tokens in the dedicated device (only known by the individual user) and a one-time password usually 4 to 6 digits [23]. This OTP is sent to the mobile device of the customer via SMS. In short, smartphones are beneficial because the customer carries them most of the time and are free of cost [24], [25]. Regardless of the popularity of SMS verification, it is recognized as the most widely adopted method for user account authentication [2]. Google and Apple introduced an SFA process for their users with the notification of new events when OTP is being delivered to the user's phone. The SMS application depends entirely on the security of the mobile phone operating system [26].

### B. Security Requirements and Attacks

The most organization today focuses on information security policies to protect their data as the Confidentiality Integrity Availability (CIA) model or Availability Integrity Confidentiality (AIC) triad Fig. 3. Confidentiality, integrity, and availability objectives should be taken together to provide information security [27]:

1) *Confidentiality*: The sensitive and personal data of the user should be secured from illegitimate access where the failure to protect the data means there is a data breach or success of the adversary to get access to the data. Impersonating, replay masquerade, spoofing, social engineering, and phishing are the most common attacks against confidentiality. While using Encryption, Quick Response code (QR), Biometrics, username, and password represent the common countermeasures techniques against these types of attacks [3].

2) *Integrity*: The security control must be able to protect the data from being modified which means the accuracy of the reverted computational outcomes upon saving or transmitting the data over networks. Insider intruders or external Man-in-the-Middle (MITM) are the common types of attacks against integrity. While encryption of the data and error detection in the transmission are common countermeasure techniques against these types of attacks [27].

3) *Availability*: To ensure the authenticated user accessibility to the resources at need or the reliability of the system uptime. Denial-of-Service (DoS) and Distributed Denial-of-Service (DDoS) are common attacks against availability. While virtual systems are the common techniques to keep availability [28].

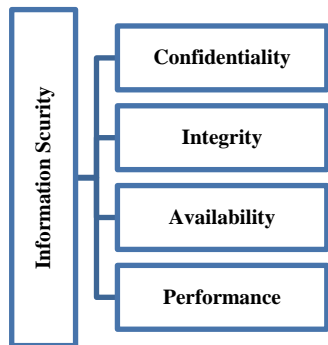


Fig. 3. CIA security requirements model.

Performance: Besides security requirements, it is important to refer to the most common requirement to implement and evaluate any authentication system performance. In addition to recognition patterns performance that its indicators can evaluate the system's ability to distinguish among the users to reduce the EER of biometric systems [29].

### III. MULTIFACTOR OF AUTHENTICATION TYPES

Authentication occurs when a user is prompted during the sign-in process to different kinds of resources like networks, devices, or apps that require the client to provide the identity. Using this identity, the client can access these resources together with authenticity proof [30] such as entering a code in the cellphone of the user or scanning a fingerprint. A basic form of authentication requires only one feature or factor, typically a password. To add another security layer, access to various resources might need over and above one factor so-called MFA whenever several factors of evidence are required [22]. There are many types of MFA as described below in Fig. 4.

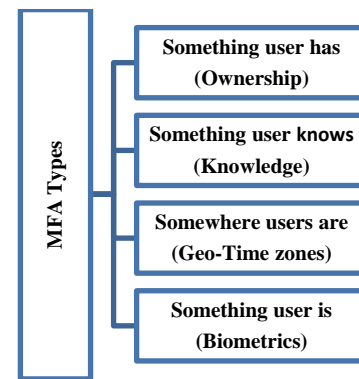


Fig. 4. MFA types.

#### A. *Something the user has (Ownership)*

Some physical things the user owns like a secret code flash drive, bank Automatic Teller Machines (ATM) card, credit card, master or visa debit cards, or hard tokens. In this regard, the Universal 2<sup>nd</sup> Factor (U2F) is an open standard that secures the transaction. It can be designed using USB or similar security technology-based Near Field Communication (NFC) devices [4], and the main drawback where the devices are stolen or lost without enough strength encryption.

#### B. *Something the user Knows (Knowledge)*

Some knowledge that only the user knows includes the password, PIN, secret question, answer, and so forth wherein the knowledge-based is the most commonly used. Herein, the user needs to reveal secret knowledge for obtaining the authentication. The password strength (complexity) represents the measure of how effectively a password can resist brute force and guessing attacks. Passwords require the use of long and random char types that are not found in the usual dictionary and might enforce the attackers to attempt all probable values [27], [31].

#### C. *Somewhere users are (Geo-Time-Zone)*

Employment of the user's geographical position acts as a location-based factor that deals with the customer's location at the login session including the physical location of the user. Upon being securely connected to the server, the user is only allowed to log in using the PIN code. While connected to the network the user might have to enter the password if necessary. This can be appropriate as access to the server for monitoring the plan depending on the time zone for the user [32], [33].

#### D. *Something user is (Biometrics)*

This type represents the biometric features that can be classified into the physical characteristics (biometrics) of the user like the fingerprints, iris, voice, facial, tongue, ears, and the vascular and micro medical biometrics like DNA, and ECG. The second type is the behavioral biometrics like typing speed, keystroke patterns, emotions, signature, gait, voice, height, gender, ethnicity, and so on are in the domain of something the user is. Biometric modal can be classified into two main types Unimodal and Multimodal [34].

1) Unimodal represents a single biometric feature.

2) Multimodal biometric systems Multimodal biometrics are developed by combining many biometrics features.

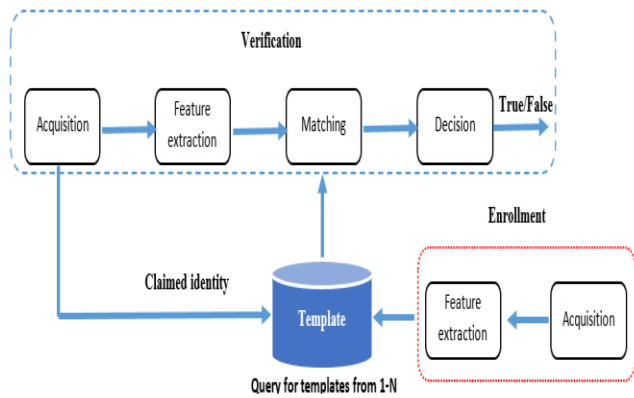


Fig. 5. Biometric system stages.

For authentication, multimodal biometrics features can be combined through fusion at different levels; sensor level, Feature-extraction level, Matching-score level, and Decision level [35]. Biometrics identity verification techniques can be used to measure the features of a user through sequences of processes Fig. 5 [36].

1) *Acquisition stage*: in a biometric system the first is to acquire the bio-data of users by sensors. For face and iris images, the sensor is typically a camera, the sensor like a camera is typically used to capture the face and iris, and the scanner is used for fingerprints and a microphone for voice. the quality of sensors has a significant impact on the performance of acquisition bio-data as environmental conditions like different sources of noises impacted the brightness, resolution of the captured image, or the depth per inch (dpi) [11] and [37].

2) *Feature extraction stage*: pre-processed to remove noise and abnormalities from the acquired bio-data to extract the bio-features for the individual ideally and uniquely. Like the fingerprint minutiae point where the position and orientation are used in the fingerprint image system to store these points during user enrollment.

3) *Matching and decision stage*: In this stage decision of accepting or rejecting is based on the matching scores comparing the features values of the stored template against the score that is generated from the enrolled stages comparing the matching score to a fixed threshold, the scores should be high to indicate the degree of similarity of the same individual (genuine matching) [38].

#### E. Biometric System Requirements

1) *Uniqueness*: The use of biometric features to identify the user's digital identity is one of the authentication methods, which must be unique, but twins can have identical features, and therefore the methodology is considered incomplete without adding other biometric factors such as fingerprints [11].

2) *Collectability*: The process of acquiring biometric properties should be easy, especially when assembled for a database system, and here it needs to be more acceptable.

3) *Performance*: As we have reviewed, the biometric system must achieve high accuracy and must be with low errors or failure to enroll, which represents the main concern of researchers and the extended challenge, Failure to Enroll (FTE), and an increase in the False Acceptance Rate (FAR) and False Rejection Rate (FRR) are among the challenges that constitute an added burden when designing multifactor authentication systems because the error is costly to the system and lead to invoke algorithms and other supporting methods when accessing the system [35].

4) *Spoofing*: There can be no guaranteed biometrics system. With the presence of various fraud methods, signatures can be forged, voices can be imitated, and fraud remains one of the most challenging attacks, as these attacks must be addressed [7].

In addition to the usual context of biometric requirements, the new requirements of the cancelable biometric systems.

1) *Non-inevitability*: original biometric template should be computationally hard to be recovered from the transformed if got compromised [11] and [29].

2) *Diversity*: should not use the identical biometric template in wide applications to prevent the reusability of the compromised templates [11] and [29].

#### F. Attacks on Biometrics System

Many attacks on the biometric system could attack at the sensor level like spoofing attacks, another at the application level like brute force attacks, and the database levels like inverse, pre-image, and dictionary attacks. Replay, presentation, MITM, and eavesdropping attacks are other types that thread the network and data transmission which can be countered through utilities time stamp or transmit the encrypted data over a secure channel [39]. On the other hand, clients' fears of bio-data breaches are one of the new challenges that the International Organization of the Red Cross has indicated, which is the refugees' fear of leaking their vital data upon receiving financial support from the organization, even in cases related to their biometric data and medical data [40].

The stolen and exposure of biometric data is not a theoretical concern, there are many existing real-world examples, in 2015 about six million fingerprints of people associated with the U.S. government sector of Office of Personnel Management breach. In 2019 of Suprema's BioStar 2 about 1 million fingerprints, as well as facial recognition information records, are stolen from the publicly accessible database [40]. In 2020 and according to Kaspersky's analysis, Kaspersky researchers referred that 37% of the servers that process and store biometric data were the target of malware attacks. [41] RSA Conference in San Francisco, In February 2019, the security expert demonstrated the success of real-time attacks like social engineering schemes, phishing SMS and emails, session hijack, and MITM attacks to intercept the traffic to circumvent the MFA[41] and [42]. In academic research, many attacks are proposed as [43] proposed an attack that used the distribution of order statistics to reverse the protected iris template of the ordinal ranking value of the

original stored iris templates. The proposed reverse-attack successes to recover greater than 95% of the template and can correctly correlate two templates of 100%. The author in [44] proposed MFA authentication of cancelable biometric hashing mechanism against the attacker is assumed to know the user's password that is bound in the bio-hash code of the face image, the scheme is based on 1-bit compressed sensing signal reconstruction by reducing the number of measurements required to acquire signals through sensing and achieved high security

### G. Biometrics System Performance Metrics

NIST SP 800-63B referred that biometrics can be utilized as important factors together with other factors (something users know). Furthermore, according to ISO/IEC 2382-37, in identification matching and authorization, the choice of the threshold value should be assigned carefully in a closed group of 0 for no match and 1 for the full match, an intruder or imposter matching score that exceeds the threshold give high FAR or the False Match Rate (FMR), while the genuine that less than the threshold gives high FRR or the False Non-Match Rate (FNMR) [42]. While the False positive (FP) denotes that exceeding the threshold of imposter scores, the False Negative (FN) sign that genuine client scores are below the threshold. Where the true positive refers to the genuine client and the true negative represents the imposter client [11], [36]. The relation of total scores can be defined in the below functions to measure the performance metrics:

- False Acceptance Rate/ False Positive Rate

$$FAR=FP/FP+FN \quad (1)$$

$$FAR = FPR = FMR * (1 - FTA) \quad (2)$$

- False Reject Rate (FRR)/ False Negative Rate(FNR)

$$FRR/FNR=FN/TP+FN \quad (3)$$

- True Acceptance Rate (TAR) or Genuine Accept Rate (GAR)

$$TAR = 1 - FRR \quad (4)$$

$$FRR=FNR= FTA + FNMR * (1 - FTA) \quad (5)$$

- True Acceptance Rate (TAR) or Genuine Accept Rate (GAR)

$$TAR = 1 - FRR \quad (6)$$

- Half Total Error Rate (HTER)

$$HTER=FNMR + FMR/2 \quad (7)$$

- Failure to Enroll rate (FTE): represents the total number of the user's failed attempts to enroll in the bio-system successfully or the number of unsuccessful attempts to enroll in the bio-system.
- The Receiver Operating Characteristic (ROC): this curve can plot the FMR on X-axis while plotting the FNMR along the Y-axis, or FAR vs FRR, or TAR vs the FAR.

- Equal Error Rate: It is another performance measurement to evaluate recognition where the FAR and FRR are equal.

$$Accuracy=TP+TN/ TP + FP + TN + FN \quad (8)$$

- Training Time: is the time of the learning algorithm to training data.
- Testing Time: this is the time of the process to test data in the learning algorithm.

### IV. AUTHENTICATION-BASED CANCELLABLE BIOMETRICS APPROACHES

The current trend of authentication research is to integrate biometric data through cancelable mechanisms that should fulfill the aforementioned cancelable biometrics requirements in subsection E of Section III. The concept of cancelability is derived from the concept of one-way functions through Cartesian, polar, functional, and hybrid transformations, but the disparity in the entries of the same user (intra-user variations) and subject to errors, and that is what we have been observed in most of the researches that attempts to improve the methods of transformation through two main approaches BioHashing, and BioEncoding [11]. [37] proposed cancelable biometrics based on a Hill cipher transformation of the biometric signals of face and palmprint as multimodal. The author in [45] proposed to use the deep Convolution Neural Network (CNN) to protect the face templates based on random projection. The first extract is the features vector from the face image with 224\*224 VGG face input. then, during the training of 15 layers 4096-dimensional output with dimensions of 1599\*4096 is projected randomly then reducing the dimension of the feature vector to 1024, the deep CNN is trained (to predict the binary code) by set neuron values to 1 if the threshold is 0.5 and 0 for else. Then the proposed deep CNN can remove the redundancies feature vector. [46] proposed cancelable biometric-based feature random projection to protect the template data against the Attack via Record Multiplicity (ARM), when the adversary may succeed to obtain multiple transformed templates from different applications to retrieve the original feature vector, which is critical to privacy and identity requirements. The basic matrix is connection with local feature slot to generate the key that is discarded after the use.

A valuable privacy-preserving research [47] proposed an authentication biometric key agreement based on cancelable biometrics. The proposed scheme integrated fuzzy commitment and Elliptic Curve Cryptography (ECC) cipher to guarantee the security of users' bio-templates against cybercrime thefts. the scheme is utilized the Random Distance Method (RDM )to generate non-invertible templates by using a random grayscale salting matrix that is added to the original feature vector values, then performing the median filter to divide each vector into two equal-size vectors to get the pseudo-biometric template.

In addition, [48] proposed protecting the transformed iris template by using ordinal ranking after XORed the user-specific string with the IrisCode string. Another proposal by [49] is to get better performance by enhancing Index First One



(IFO) hashing for iris templates. The binary confidence matrix considered the variation in noisy iris Biometric Template Protection (BTP) systems. the Fully Connected Architecture (FCA) and Bilinear Architecture (BLA) are used by [50] to hash a binary vector template. The proposed framework based on Deep Neural Networks (DNN) was tested on 50 subjects only. The author in [51] proposed a Multi-Instance Cancelable iris authentication Deep Learning (MICBTDL) and used a CNN (triplet loss) and trained to differentiate a positive image from a negative on IITD and MMU iris dataset images. Both [52] and [53] proposed to encrypt the Iris Codes using classical cryptography algorithms.

Table I presents the summarization regarding CB systems and provides an overall idea about the recent direction of authentication-based CB. Ten recent works have been summarized based on four categories which are: “purpose of the research”, then “proposed methods”, “Bio-Feature”, “Bio-Dataset” and finally “**Performance**” measurement of the research that has been used in the experiment.

TABLE I. SUMMARY ANALYSIS OF THE RECENT CB SCHEMES

Ref.	CB Scheme	Bio-Features	Bio-Dataset	Performance
[37]	Hill Cipher	Face Palmprint	Face={ORL, Indian Face, Yale} Palmprint={PolyU, CASIA}	EER
[49]	Confidence matrix	Iris	CASIA Iris v4-interval	EER
[45]	Random projection	Face	CMU-PIE, FEI, Color-FERET	GAR
[46]	Feature-adaptive random projection	Fingerprint	FVC2002 DB1-DB3 and FVC2004 DB2	EER
[52]	Encryption (3DES+Twofish)	Iris	CASIA-IrisV3	GAR
[47]	Random distance method (RDM)	Fingerprint Face Iris	Fingerprint={FVC2006} Face={CASIA-Face V5} IRIS={IRIS(LWIR)}	FAR, EER, ROC communication computational
[53]	Encryption (AES)	Iris	CASIA-IrisV3	EER
[48]	Local rank	Iris	CASIA Iris v3-interval	EER
[50]	Deep neural networks Integration	Face Iris	Face={ Casia-Webface} Iris={CASIA-Iris-Thousand}	GAR
[51]	MICBTDL	Iris	IITD+ MMU	EER

### V. NON-BIOMETRICS AUTHENTICATION METHODS

Previous literature demonstrates authentication-based cancelable biometrics to account for authentication mechanisms and the current methods to enhance the pattern recognition of biometric features.

In this section, non-based biometrics authentication methods have been summarized according to the methods for generation and distribution. Also, tabular evaluation matrices have been summarized in each section. These evaluation matrices provide a structured technique for the outcome of the metrics and the criteria of the proposed methods. In OTP Authentication Procedure The server customer or application (Object/Service) as a client has to register an online account and provide the mobile number and other required information

during the registration process so that online login to the bank server (as an example) is successful (Fig. 6).

1) The server exchanges the OTP code with the registered mobile number, email, or application, and the code is used as a second layer of the authentication mechanism for any transaction achieved through the following steps [10], [54].

2) The client requests a transaction to connect with the server using his credentials username and password.

3) The server checks the client validity and if successful then initiates the OTP generation algorithm.

4) The output of step ii should be encrypted using cryptography algorithms (Symmetric or Asymmetric) or hashing then forward to the client.

5) The client's application decrypts the obtained OTP if it is encrypted on the server side.

6) The client's application gets the OTP code in step 4 and enters it manually or automatically in the text box of the bank application.

7) The server receives the code from the client then checks the comparison condition if success then completes the transaction during the timer period. The server should acknowledge the client for the verification of the transaction status (success or failure).

Steps 2 and 3 being optional are recommended according to NIST special publication 800-63B privacy requirements [55] against various attacks such as birthday, offline/online guessing, impersonation, rainbow, a man in the middle, key-recovery, and collision types.

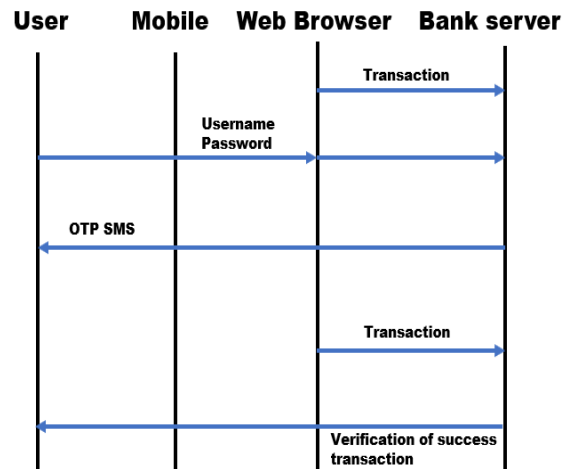


Fig. 6. One-time passcode workflow.

### B. One-Time Passcode Generation Methods

The first attempt to overcome the weaknesses of finite OTP generation for authentication was by [56] using a hush chain. The author in [57] proposed a model to generate OTP. The model addresses the weaknesses of the Lamport method when finite generation OTP and used the hash chain function. However, the bottleneck of the proposed model is the server generates an unlimited OTP code for each client login. The author in [58] proposed a model to overcome the weaknesses of the OTP model of Lamport using a Merkle tree an L-divided

tree (with the binary tree), and the hash chain. The author illustrated the proposed model examples without implementation. A model was proposed [59] for generating the OTP using the following steps. The first step was based on the certification authority's client login details user name and password and converting it to 8 bits. The second step converted the message's length to 8 bits. In the third step, a binary code is converted to decimal base (CB2D) was used to convert 16 bits ( $2 \times 8$  bits) to base decimal, the conversion result is two 3 decimal digits. Fourth step used 3 Vedic multipliers to get the OTP multiplied by two decimal digits. The proposed OTP generation method used the user name and the password credentials without encryption.

A novel method was introduced [60] for improving the verification in mobile banking using lightweight and solid OTP generators. The OTP generation model is based on three client factors (6 digits of the last transaction date DD: MM: YY. The last transaction amounted to 16 digits, wherein three login characters of the password were selected. The client must have an SD card containing initialization vectors (IV) set of 256 bits, AES-256-XTS encrypted random keys, each key of 256 bits was divided into two portions, each one containing 128 bits on the user's side and the other half on the server side. Then it was XOR'ed followed by concatenating and hashing together via the SHA 256 for using the AES-256-CBC to encrypt the OTP. Afterward, it was sent to the client via the internet to the mobile device of the client to make a comparison of the transaction is succeeded. All the used 256 keys were stored in the hardware component SD Card with the bank client. Later, the OTP was sent only through the internet only when the client application initiates a session of user online login. OTP length is 64 digits only, without considering using case-sensitive characters like uppercase letters or symbols, charset sizes are medium (0-9, a-z).

An OTP generation scheme is proposed by [31] for authentication purposes. The first step encrypts a randomly

selected image stored in the server database using Bitwise Masking Alternate Sequence (BWMAS). A two-step authentication model was proposed to generate the OTP on many intermediate operations. Then in the second stage, OTP was sent by e-mail to the client. The researchers in [61] proposed an algorithm for the implementation of secure OTP generation. OTP was a mixture of the present time value of 3 digits from the login time (HH: MM: SS) to shuffle the hex and octal which is obtained via the back-end server-generated value random operation. This value is 32-bits in size, producing an 8-digit of number combination. Fixed length OTP of 8 digits only is obtained without including the small or capital letters as well as symbols, generating a complexity of  $232 = 4294967296$  combinations.

Table II elucidates the OTP generation using various schemes [59] [60], [31], and [61] including the proposed methods, initial data of OTP generation, encryption mechanisms, hash functions, delivery methods types such as email, SMS, Internet, and the authentication protocol.

### C. One-Time Passcode Distribution Methods

First, the client has to enroll in the first phase at the server for receiving the OTPs. In addition, the client must declare the procedure of distributing the OTP via his/her email or SMS, or some apps [57]. An approach was proposed [62] for online e-banking authentication using OTP. This approach provided a mechanism for the generation of an infinite and forward OTP utilizing SHA3 and SHA2 followed by the dynamical concatenation understandable by humans. An inimitable verification scheme was employed with a distinct initial seed for generating several OTPs on the users' mobile devices. This method showed superior performance compared to the existing authorization techniques. This is due to the eradication, during the authorization process, of cellular networks. The findings that are critical for online transactions have shown a high degree of success and efficacy in the verification and authorization.

TABLE II. SUMMARY ANALYSIS OF DIFFERENT EXISTING ONE-TIME PASSCODE GENERATION STUDIES

Ref.	Methods	Initial data	Cryptography	Hash Functions	Media	Authentication
[59]	CBCD and 3 *3 Vedic Multiplier	Client's User name and Password	No	No	Internet	MFA
[60]	The OTP generation model is based on client factors and then encrypted and hashed them and forwarded to the server-side as a 64digits length	selected three client factors: last transaction date DD: MM: YY, Last transaction amount to 6 bits, three login password characters	AES-256	SHA2-256	SMS	MFA
[31]	OTP is generated by applying Bit-wise Masking and Alternate Sequence (BWMAS) on selecting a random image stored at the server database Provided by Clients to generate numeric OTP.	Biometric Fingerprint	No	No	Email/SMS	MFA
[61]	Server-generated random value OTP would be a combination of the current time value of 3 digits from the login time to shuffle the hex and octal to that combined value. This value is 32-bit in size, which produces an 8-digit number combination	Current Time of login session HH:MM: SS	No	No	SMS	MFA

The OTP is sent via cellular data to the client, without using encryption methods to protect data privacy. Generally, the methods for the OTPs distribution are multiple and have their own merits. The business organization looks for a method of OTP verification distribution, especially via SMS and email. The OTP-based SMS is usually utilized for the users' login and password resetting. Several organizations particularly the financial sectors send the OTPs via SMS and email as an additional confirmation step for the users. The following modes are used:

- 1) *Email-Based*: The OTP can be delivered to the client through his/her Email [63].
- 2) *SMS Based*: The OTP can be delivered to the client through his mobile phone over-the-air SMS or delivered to an OTP application [63].
- 3) *Push notifications*: These can be sent to the client's registered smartphone for out-of-band authentication which could not be intercepted at the point of password entry when the client approves the request. Then, the authentication application informs the server that the client is confirmed. The authentication-based push notification provides a higher level of security since it addresses the most important SMS and email shortcomings, such as vulnerability to replay attacks. It is even faster than typing in a passcode [64], [65].
- 4) *Hard copies*: Some banks provide clients with hard-printed OTPs for each transaction [63].

Many methods are used to distribute an SMS or email wherein each has its weaknesses. The OTP distribution schemes are simpler for the companies because of the abundance of the SMTP to SMS channels. No setting is required for the clients wherein they just request a code to log in. The ease of administration related to the delivery of OTPs via SMS and email is frequently utilized, enabling temporary access. Nonetheless, OTP is MFA which is something that the user knows, and the mobile devices (phones, computers, smart-watches, tablets, and so forth) something that the user is not essentially true.

Using email and SMS delivery protocols (something one has) the OTP is sent to the user. Presently, several phone numbers can be cloned which affects the authentication of the client identity [66], [67]. In this regard, email phishing, spear-phishing, and scammers can develop ways that entice the

clients to enter a username or password and follow a suspicious link or download malware applications. Thus, visiting certain authentic-looking websites permits the adversary to steal the credentials or gain control of the users' devices [62], [63].

The OTP overcomes the shortcomings of the SFA like traditional password because it's not reusable and it can be distributed through the Out of Bound (OOB) channel as well as not vulnerable to replay attacks. Therefore, it's more secure to transmit an OTP over a push notification. Due to the rapid adoption and fast growth of push authentication, it offers a low-cost, easy-to-use, and secure substitution for email and SMS distribution methods [68].

Table III shows the comparison of main property qualities such as security, message size, feedback possibilities, response time, and cost among SMS, email, and push notifications. It is argued that push notification is the best method that could be used for OTP distribution due to the low cost, high speed, and more secure than other methods of OTP distribution. While in [69] the authors proposed a model to improve the OTP distribution security using Elliptic Curve Cryptography (ECC) and Iris for key generation. The model encrypted the generated OTP and sends it as ciphertext, not as plain text. The use of ECC with the iris code-based public key could encrypt the OTP successfully and send it through email. The author in [70] utilized lightweight cryptography and text steganography to encrypt and hide OTP to send the stego text as an SMS to the mobile application of the client. This method was used to protect the delivery of OTP SMS where the OTP was encrypted and steganography ciphering was exploited for hiding it in a standard SMS. This process used the Date of Birth (DOB) of the customer with the secret four digits of PIN only as a key to encrypt and decrypt the OTP. The author in [71] proposed a mechanism for protecting OTP delivery from many types of attacks, especially smartphone Trojans. A virtual dedicated channel was used to secure SMS-based OTPs against cell phone Trojans stealing SMS. By assigning a port to transmit and receive OTP via TCP/ UDP from the service provider, and protect the mobile application's storage files internally.

Table IV enlists the secure end-to-end distribution of OTP [69], [70], [71] based on many categories such as proposed methods, encryption mechanisms, hash functions, delivery methods types (for example email, SMS, internet), and authentication protocol.

TABLE III. COMPARISON OF MAIN QUALITIES IN ONE-TIME PASSCODE DISTRIBUTION STUDIES

Parameter	SMS	Email	Push Notification
Security [20]	Medium	Low	High
Message size [72]	Limited (140 Bytes)	Long	Moderate (according to app settings).
Feedback Possibilities [73], [74]	Yes (If the sender is Known)	Yes (If the sender is Known)	Yes (according to app settings).
Response Time [73], [74]	Medium	Low	High
Usability of intuitive user interfaces[75]	Medium	Low	High
Cost [76]	High	Low	Low

TABLE IV. SUMMARY ANALYSIS OF DIFFERENT ONE-TIME PASSCODE DISTRIBUTION STUDIES

Ref.	Methods	Initial data	Cryptography	Hash Functions	Media	Authentication
[60]	Traditional OTP Generation and Distribution	Numeric (0-9)	No	No	SMS	MFA
[69]	Secure OTP distribution	Iris of client	Elliptic Curve Cryptography (ECC)	No	Email	MFA
[70]	Secure OTP distribution	Client date of Birth DOB	Steganography fixed key, lightweight Feistel cipher	No	Internet	MFA
[71]	OTP distribution	Assign a port to receive OTP in the user client's mobile phone	No	No	Use a virtual dedicated channel from ISP	MFA

## VI. MAIN CHALLENGES

The combination of new solutions is always challenging for both developers and managers for the implementation of a strong identity in usability and resistance against known attacks [5]. So it is worth mentioning that issues for authentication must be tackled regarding the other digital identity management mechanisms in authentication.

### A. This Paper has Concerned the MFA Regarding the Cancelable Biometrics Approach and Shows the following Challenges

1) Most of the biometric databases vary in type and version while low accuracy when in real use, beside train biometric recognition systems through machine learning or deep learning results in varying accuracy metrics.

2) Choosing a set of biometric features to design appropriate authentication systems is a critical challenge in designing MFA based on multimodal cancelable biometrics.

3) Low user adaptability due to additional hardware requirements for sensing and processing persists despite smartphone applications.

4) Concerns increase for biometric data leakage despite cancelable biometric requirements.

5) Difficulties in pattern recognition due to the noise, or the variations of large intraclass of transformed templates.

### B. In the Second MFA Approach (Non-based Biometrics)

The main challenge of non-based biometrics is explained in [77] where the impossibility of distinguishing the correlation between the ideal-random and pseudo-random sequences in the absence of unlimited computational capacity. While in some cases the clients may be unable to access the OTP codes offline or without a network (for example in an airplane during international journeys) [78].

In addition to the growth and success of session hijacking attacks, research on both approaches did not concern this type of attack. The surveyed studies did not concern with its proposing of authorization and accounting mechanisms, most of them were simulated without paying attention to the impacts of decision-making (only pattern recognition enhancement) by the identity provider based on the backend database and customer account data as well as the target resource, because the accessing to a particular resource must direct the matcher according to the conditions of access controls policies to direct or redirect the accessing of one or several resources.

Circle of verification and loop of authentication, this means without a controller or police (based biometric or non-based biometric) in algorithm design, the enrollment and verification of the client identity will repeat and will call or invoke the system resources (objects, algorithms, procedures) that require access from one moment to another and between one object and another, and this is costly due to consume the two parties of the communication system resources. While the researchers in [79] pointed threat of involving third-party authentication, especially when this requires the customer's biometric data. Also, the research [79] indicated one of the main challenges today is to ensure that the customer's biometric data is destroyed and not used again (no copy of such biometric data on the server side), and to confirm that to the customer.

## VII. CONCLUSION AND FUTURE DIRECTORIES

The digital identities and MFA are intensively combined in the context of advanced information communication technologies. This paper comprehensively reviewed various MFA in two approaches biometric and non-biometric mechanisms besides pointing out many standard protocols. Effective non-biometrics MFA schemes are required to combine many factors and protocols to secure the data distribution in an end-to-end. Cancelable biometrics approaches are proposed to transform the templates like iris to be difficult to obtain the original traits but still suffer from difficulties to maintain the tradeoff between security and performance while supporting effective biometrics recognition as shown in the literature. Also, we presented different performance measures used for cancelable biometrics in both identification and verification. Now, we also draw attention to future directions in cancelable biometrics. The readers may look for propose new mechanisms to transform the templates and maintain the recognition performance while protecting the privacy of biometric data. The employment of deep learning in CB requires a large number of training samples. In some cases there is only a single trait for each subject is available that imposes limits and constraints on the training. However, the future of MFA appears boundless with absolutely secured data exchange over the internet and cellular. Concisely, the ongoing extensive research activities in the MFA are likely to transform many facets of the forthcoming evolution and continue to be more secure.

## REFERENCES

- [1] F. Sinigaglia, R. Carbone, G. Costa, and N. Zannone, "A survey on multi-factor authentication for online banking in the wild," *Comput. Secur.*, vol. 95, 2020, doi: 10.1016/j.cose.2020.101745.

- [2] F. Liébana-Cabanillas, I. R. de Luna, and F. Montoro-Riosa, "Intention to use new mobile payment systems: A comparative analysis of SMS and NFC payments," *Econ. Res. Istraz.*, vol. 30, no. 1, pp. 892–910, 2017, doi: 10.1080/1331677X.2017.1305784.
- [3] T. Mehraj, M. A. Sheheryar, S. A. Lone, and A. H. Mir, "A critical insight into the identity authentication systems on smartphones," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 13, no. 3, pp. 982–989, 2019, doi: 10.11591/ijeecs.v13.i3.pp982-989.
- [4] P. A. Grassi, M. E. Garcia, and J. L. Fenton, "Digital identity guidelines: revision 3," *Digit. Identity Guidel.*, vol. 58, no. 2, pp. 130–137, 2020, [Online]. Available: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-63-3.pdf>.
- [5] A. Ometov, S. Bezzateev, N. Mäkitalo, S. Andreev, T. Mikkonen, and Y. Koucheryavy, "Multi-Factor Authentication: A Survey," *Cryptography*, vol. 2, no. 1, p. 1, 2018, doi: 10.3390/cryptography2010001.
- [6] A. Ometov and S. Bezzateev, "Multi-factor authentication: A survey and challenges in V2X applications," *Int. Congr. Ultra Mod. Telecommun. Control Syst. Work.*, vol. 2017-Novem, pp. 129–136, 2017, doi: 10.1109/ICUMT.2017.8255200.
- [7] I. Stylios, S. Kokolakis, O. Thanou, and S. Chatzis, "Behavioral biometrics & continuous user authentication on mobile devices: A survey," *Inf. Fusion*, vol. 66, no. February 2020, pp. 76–99, 2021, doi: 10.1016/j.inffus.2020.08.021.
- [8] S. Das, B. Wang, Z. Tingle, and L. Jean Camp, "Evaluating user perception of multi-factor authentication a systematic review," *arXiv*, 2019.
- [9] G. Ali, M. A. Dida, and A. E. Sam, "Two-factor authentication scheme for mobile money: A review of threat models and countermeasures," *Futur. Internet*, vol. 12, no. 10, pp. 1–27, 2020, doi: 10.3390/fi12100160.
- [10] H. Shahriar, T. Klintic, and V. Clincy, "Mobile Phishing Attacks and Mitigation Techniques," *J. Inf. Secur.*, vol. 06, no. 03, pp. 206–212, 2015, doi: 10.4236/jis.2015.63021.
- [11] Manisha and N. Kumar, "Cancelable Biometrics: a comprehensive survey," *Artif. Intell. Rev.*, vol. 53, no. 5, pp. 3403–3446, 2020, doi: 10.1007/s10462-019-09767-8.
- [12] R. Y. Zakari, A. Suleiman, Z. K. Lawal, and N. Abdulrazak, "A Review of SMS Security Using Hybrid Cryptography and Use in Mobile Money System," *Am. J. Comput. Sci. Eng.*, vol. 2, no. 6, pp. 53–62, 2015.
- [13] M. Noman Riaz and A. Ikram, "Development of a Secure SMS Application using Advanced Encryption Standard (AES) on Android Platform," *Int. J. Math. Sci. Comput.*, vol. 4, no. 2, pp. 34–48, 2018, doi: 10.5815/ijmsc.2018.02.04.
- [14] A. Saha and S. Sanyal, "Survey of Strong Authentication Approaches for Mobile Proximity and Remote Wallet Applications - Challenges and Evolution," *Int. J. Comput. Appl.*, vol. 108, no. 8, pp. 10–15, 2014, doi: 10.5120/18930-0319.
- [15] I. Amazon Web Services, "AWS Identity and Access Management: Benutzerhandbuch," p. 2106, 2020.
- [16] P. Pandey and T. N. Nisha, "Challenges in Single Sign-On," *J. Phys. Conf. Ser.*, vol. 1964, no. 4, 2021, doi: 10.1088/1742-6596/1964/4/042016.
- [17] E. B. Barker, M. Smid, and D. Branstad, "NIST Special Publication 800-152 - A Profile for U. S. Federal Cryptographic Key Management Systems," *NIST Spec. Publ.*, 2015, [Online]. Available: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-152.pdf>.
- [18] E. Barker, A. Roginsky, G. Locke, and P. Gallagher, "Transitions: Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths," *NIST Spec. Publ.*, vol. 1, no. January, pp. 800–131, 2011.
- [19] E. Belmekki, B. Raouyane, A. Belmekki, and M. Bellafkih, "Secure SIP signalling service in IMS network," 2014 9th Int. Conf. Intell. Syst. Theor. Appl. SITA 2014, no. May, 2014, doi: 10.1109/SITA.2014.6847291.
- [20] P. Doucek, L. Pavlíček, J. Sedláček, and L. Nedomová, "Adaptation of password strength estimators to a non-english environment—the Czech experience," *Comput. Secur.*, vol. 95, 2020, doi: 10.1016/j.cose.2020.101757.
- [21] A. S and K. S. Anil Kumar, "Security and performance enhancement of fingerprint biometric template using symmetric hashing," *Comput. Secur.*, vol. 90, 2020, doi: 10.1016/j.cose.2020.101714.
- [22] Shally and G. S. Aujla, "A review of one time password mobile verification," *Int. J. Comput. Sci. Eng. Inf. Technol. Res.*, vol. 4, no. 3, pp. 113–118, 2014.
- [23] S. Ma et al., "An empirical study of SMS one-time password authentication in android apps," *ACM Int. Conf. Proceeding Ser.*, pp. 339–354, 2019, doi: 10.1145/3359789.3359828.
- [24] A. Dmitrienko, C. Liebchen, C. Rossow, and A.-R. Sadeghi, "Security analysis of mobile two-factor authentication schemes," *Intel Technol. J.*, vol. 18, no. 4, pp. 138–161, 2014, [Online]. Available: <http://ezproxy.library.capella.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=97377858&site=ehost-live&scope=site>.
- [25] N. R. Dive, M. D. Likhar, N. A. Ughade, S. S. Chune, and M. Khonde, "Survey Of Graphical Password Authentication Techniques," no. 3, pp. 145–152, 2016.
- [26] H. S. Alsaiani, "Graphical One-Time Password Authentication," 2016.
- [27] M. Botacin, F. Ceschin, P. de Geus, and A. Grégio, "We need to talk about antiviruses: challenges & pitfalls of av evaluations," *Comput. Secur.*, vol. 95, 2020, doi: 10.1016/j.cose.2020.101859.
- [28] B. A. Buhari, A. A. Obiniyi, K. Sunday, and S. Shehu, "Performance Evaluation of Symmetric Data Encryption Algorithms: AES and Blowfish," *Saudi J. Eng. Technol.*, vol. 04, no. 10, pp. 407–414, 2019, doi: 10.36348/sjeat.2019.v04i10.002.
- [29] S. Mansfield-Devine, "The ever-changing face of phishing," *Comput. Fraud Secur.*, vol. 2018, no. 11, pp. 17–19, 2018, doi: 10.1016/S1361-3723(18)30111-8.
- [30] Y. Shah, V. Choyi, A. U. Schmidt, and L. Subramanian, "Multi-factor authentication as a service," *Proc. - 2015 3rd IEEE Int. Conf. Mob. Cloud Comput. Serv. Eng. MobileCloud 2015*, no. March, pp. 144–150, 2015, doi: 10.1109/MobileCloud.2015.35.
- [31] J. Bhaumik and I. Chakrabarti, *Communication , Devices , and Computing*. 2017.
- [32] M. Conti, N. Dragoni, and V. Lesyk, "A Survey of Man in the Middle Attacks," *IEEE Commun. Surv. Tutorials*, vol. 18, no. 3, pp. 2027–2051, 2016, doi: 10.1109/COMST.2016.2548426.
- [33] E. De Cristofaro, H. Du, J. Freudiger, and G. Norcie, "A Comparative Usability Study of Two-Factor Authentication," no. March, 2014, doi: 10.14722/usec.2014.23025.
- [34] A. Lumini and L. Nanni, "Overview of the combination of biometric matchers," *Inf. Fusion*, vol. 33, pp. 71–85, 2017, doi: 10.1016/j.inffus.2016.05.003.
- [35] V. Conti, C. Militello, F. Sorbello, and S. Vitabile, "A frequency-based approach for features fusion in fingerprint and iris multimodal biometric identification systems," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 40, no. 4, pp. 384–395, 2010, doi: 10.1109/TSMCC.2010.2045374.
- [36] R. Dwivedi and S. Dey, "A non-invertible cancelable fingerprint template generation based on ridge feature transformation," vol. 4, pp. 1–17, 2018, [Online]. Available: <http://arxiv.org/abs/1805.10853>.
- [37] H. Kaur and P. Khanna, "Non-invertible biometric encryption to generate cancelable biometric templates," *Lect. Notes Eng. Comput. Sci.*, vol. 1, no. 1, pp. 432–435, 2017.
- [38] M. A. Al-Shareeda, M. Anbar, S. Manickam, and I. H. Hasbullah, "An efficient identity-based conditional privacy-preserving authentication scheme for secure communication in a vehicular ad hoc network," *Symmetry (Basel)*, vol. 12, no. 10, pp. 1–25, 2020, doi: 10.3390/sym12101687.
- [39] A. O. Alaswad, A. H. Montaser, and F. E. Mohamad, "Vulnerabilities of Biometric Authentication Threats and Countermeasures," *Int. J. Inf. Comput. Technol.*, vol. 4, no. 10, pp. 947–958, 2014, [Online]. Available: <http://www.irphouse.com>.
- [40] B. Hayes(ICRC), "Facilitating innovation, ensuring protection: the ICRC Biometrics Policy - Humanitarian Law & Policy Blog | Humanitarian Law & Policy Blog," 2019. <https://blogs.icrc.org/law-and->

- policy/2019/10/18/innovation-protection-icrc-biometrics-policy/  
(accessed Aug. 23, 2021).
- [41] “Five ways to hack MFA and the FBI’s mitigation strategy.”
- [42] T. Seals, “ThreatList: A Third of Biometric Systems Targeted by Malware in Q3 | Threatpost,” 2019. <https://threatpost.com/threatlist-a-third-of-biometric-systems-targeted-by-malware-in-q3/150778/> (accessed Aug. 23, 2021).
- [43] O. Ouda, “On the Practicality of Local Ranking-Based Cancelable Iris Recognition,” *IEEE Access*, vol. 9, pp. 86392–86403, 2021, doi: 10.1109/access.2021.3089078.
- [44] B. Topcu, C. Karabat, M. Azadmanesh, and H. Erdogan, “Practical security and privacy attacks against biometric hashing using sparse recovery,” *EURASIP J. Adv. Signal Process.*, pp. 1–20, 2016, doi: 10.1186/s13634-016-0396-1.
- [45] A. K. Jindal, S. Rao Chalamala, and S. K. Jami, “Securing Face Templates using Deep Convolutional Neural Network and Random Projection,” 2019 IEEE Int. Conf. Consum. Electron. ICCE 2019, 2019, doi: 10.1109/ICCE.2019.8662094.
- [46] W. Yang, S. Wang, M. Shahzad, and W. Zhou, “A cancelable biometric authentication system based on feature-adaptive random projection,” *J. Inf. Secur. Appl.*, vol. 58, p. 102704, 2021, doi: 10.1016/j.jisa.2020.102704.
- [47] L. Wu, L. Meng, S. Zhao, X. Wei, H. Wang, and L. Wu, “Privacy-preserving Cancelable Biometric Authentication Based on RDM and ECC,” *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3092018.
- [48] D. Zhao, S. Fang, J. Xiang, J. Tian, and S. Xiong, “Iris Template Protection Based on Local Ranking,” *Secur. Commun. Networks*, vol. 2018, 2018, doi: 10.1155/2018/4519548.
- [49] T. Y. Chai, B. M. Goi, and W. S. Yap, “Towards better performance for protected iris biometric system with confidence matrix,” *Symmetry (Basel)*, vol. 13, no. 5, 2021, doi: 10.3390/sym13050910.
- [50] V. Talreja, M. C. Valenti, and N. M. Nasrabadi, “Multibiometric secure system based on deep learning,” 2017 IEEE Glob. Conf. Signal Inf. Process. Glob. 2017 - Proc., vol. 2018-Janua, pp. 298–302, 2018, doi: 10.1109/GlobalSIP.2017.8308652.
- [51] M. Sandhya, M. K. Morampudi, I. Pruthweraaj, and P. S. Garepally, “Multi-instance cancelable iris authentication system using triplet loss for deep learning models,” *Vis. Comput.*, 2022, doi: 10.1007/s00371-022-02429-x.
- [52] O. C. Abikoye, U. A. Ojo, J. B. Awotunde, and R. O. Ogundokun, “A safe and secured iris template using steganography and cryptography,” 2020.
- [53] A. A. Asaker, Z. F. Elsharkawy, S. Nassar, N. Ayad, O. Zahran, and F. E. Abd El-Samie, “A novel cancellable Iris template generation based on salting approach,” *Multimed. Tools Appl.*, vol. 80, no. 3, pp. 3703–3727, 2021, doi: 10.1007/s11042-020-08663-6.
- [54] I. Standard, “INTERNATIONAL STANDARD ISO / IEC Information technology — JPEG 2000,” vol. 2007, 2007.
- [55] J. L. Fenton et al., “Digital identity guidelines: Authentication and Lifecycle Management,” *NIST Spec. Publ. 800-63B*, pp. 2–79, 2017, [Online]. Available: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-63b.pdf>.
- [56] L. Lamport, “Password Authentication with Insecure Communication,” *Commun. ACM*, vol. 24, no. 11, pp. 770–772, 1981, doi: 10.1145/358790.358797.
- [57] C. S. Park, “One-time password based on hash chain without shared secret and re-registration,” *Comput. Secur.*, vol. 75, pp. 138–146, 2018, doi: 10.1016/j.cose.2018.02.010.
- [58] Y. Suga, “An Extended Lamport-Like One-Time Password Scheme and its Applications,” 2018 IEEE Int. Conf. Consum. Electron. - Asia, ICCE-Asia 2018, pp. 6–9, 2018, doi: 10.1109/ICCE-ASIA.2018.8552134.
- [59] S. P. Shyry, M. Mahithasree, and M. Saranya, “Implementation of One Time Password by 3\*3 Vedic Multiplier,” 2nd Int. Conf. Comput. Commun. Signal Process. Spec. Focus Technol. Innov. Smart Environ. ICCSP 2018, no. Iccsp, 2018, doi: 10.1109/ICCSP.2018.8452861.
- [60] H. S. Elganzoury, A. A. Abdelhafez, and A. A. Hegazy, “2018 , 35 th NATIONAL RADIO SCIENCE CONFERENCE A New Secure One-Time Password Algorithm for Mobile Applications 2018 , 35 th NATIONAL RADIO SCIENCE CONFERENCE,” no. Nrsc, pp. 249–257, 2018.
- [61] “Gosavi, S. S., & Shyam, G. K. (2020). A Novel Approach of OTP Generation Using Time-Based OTP and Randomization Techniques. In *Data Science and Security* (pp. 159-167). Springer, Singapore.,” 2020.
- [62] S. Hussain, B. U. I. Khan, F. Anwar, and R. F. Olanrewaju, “Secure Annihilation of Out-of-Band Authorization for Online Transactions,” *Indian J. Sci. Technol.*, vol. 11, no. 5, pp. 1–9, 2018, doi: 10.17485/ijst/2018/v11i5/121107.
- [63] M. M. Hashim, M. Shafry, and M. Rahim, “a Review and Open Issues of Diverse Text,” vol. 96, no. 17, pp. 5819–5840, 2018.
- [64] M. Sahin, A. Francillon, P. Gupta, and M. Ahamad, “SoK: Fraud in Telephony Networks,” *Proc. - 2nd IEEE Eur. Symp. Secur. Privacy, EuroSP 2017*, pp. 235–240, 2017, doi: 10.1109/EuroSP.2017.40.
- [65] S. M. Seth and R. Mishra, “Comparative Analysis Of Encryption Algorithms For Data Communication,” *Ijst*, vol. 2, no. 2, pp. 292–294, 2011.
- [66] B. Reaves, N. Scaife, D. Tian, L. Blue, P. Traynor, and K. R. B. Butler, “Sending Out an SMS: Characterizing the Security of the SMS Ecosystem with Public Gateways,” *Proc. - 2016 IEEE Symp. Secur. Privacy, SP 2016*, pp. 339–356, 2016, doi: 10.1109/SP.2016.28.
- [67] A. R. L. Reyes, E. D. Festijo, and R. P. Medina, “Enhanced multi-factor out-of-band authentication en route to securing SMS-based OTP,” *Int. J. Eng. Technol. Innov.*, vol. 9, no. 2, pp. 145–154, 2019.
- [68] P. A. Grassi et al., “NIST Special Publication 800-63B,” 2019, Accessed: Jan. 23, 2021. [Online]. Available: <https://pages.nist.gov/800-63-3/sp800-63b.html>.
- [69] “Mahto, D., & Yadav, D. K. (2016). Security improvement of one-time password using crypto-biometric model. In *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics* (pp. 347-353). Springer, New Delhi.,” 2016.
- [70] S. Chandra and S. Tai, *Advances in Intelligent Systems and Computing 828 Proceedings of the 2nd International Conference on Data Engineering and Communication Technology*, vol. 2, 2017.
- [71] S. Paper, P. Stewin, and J. Seifert, “SMS-Based One-Time Passwords : Attacks and Defense,” pp. 150–159, 2013.
- [72] A. Hernández-Reyes, G. Molina-Recio, R. Molina-Luque, M. Romero-Saldaña, F. Cámara-Martos, and R. Moreno-Rojas, “Effectiveness of PUSH notifications from a mobile app for improving the body composition of overweight or obese women: A protocol of a three-Armed randomized controlled trial,” *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, pp. 1–10, 2020, doi: 10.1186/s12911-020-1058-7.
- [73] L. G. Morrison et al., “The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: An exploratory trial,” *PLoS One*, vol. 12, no. 1, pp. 1–15, 2017, doi: 10.1371/journal.pone.0169162.
- [74] A. Wohllebe, “Consumer Acceptance of App Push Notifications: Systematic Review on the Influence of Frequency,” *Int. J. Interact. Mob. Technol.*, vol. 14, no. 13, p. 36, 2020, doi: 10.3991/ijim.v14i13.14563.
- [75] V. Picchio, V. Cammisotto, F. Pagano, R. Carnevale, and I. Chimenti, “We are IntechOpen , the world ’ s leading publisher of Open Access books Built by scientists , for scientists TOP 1 %,” *Intechopen*, no. Cell Interaction-Regulation of Immune Responses, Disease Development and Management Strategies, pp. 1–15, 2020, [Online]. Available: <https://www.intechopen.com/books/advanced-biometric-technologies/liveness-detection-in-biometrics>.
- [76] S. M. Abdulhamid et al., “A Review on Mobile SMS Spam Filtering Techniques,” *IEEE Access*, vol. 5, pp. 15650–15666, 2017, doi: 10.1109/ACCESS.2017.2666785.
- [77] H. Kim, J. Han, C. Park, and O. Yi, “Analysis of vulnerabilities that can occur when generating one-time password,” *Appl. Sci.*, vol. 10, no. 8, 2020, doi: 10.3390/APP10082961.
- [78] T. S. Mohamed, “Security of Multifactor Authentication Model to Improve Authentication Systems,” *Inf. Knowl. Manag.*, vol. 4, no. 6, pp. 81–87, 2014.

- [79] S. Barra, K. K. R. Choo, M. Nappi, A. Castiglione, F. Narducci, and R. Ranjan, "Biometrics-as-a-service: Cloud-based technology, systems, and applications," *IEEE Cloud Comput.*, vol. 5, no. 4, pp. 33–37, 2018, doi: 10.1109/MCC.2018.043221012.

# Arabic Dialogue Processing and Act Classification using Support Vector Machine

Abraheem Mohammed Sulayman Alsubayhay, Md Sah Hj Salam, Farhan Bin Mohamed  
School of Computing-Faculty of Engineering, Universiti Teknologi Malaysia (UTM), Johor, Malaysia

**Abstract**—Text classification is the technique of grouping documents according to their content into classes and groups. As a result of the vast amount of textual material available online, this procedure is becoming increasingly crucial. The primary challenge in text categorization is enhancing classification accuracy. This role is receiving more attention due to its importance in the development of these systems and the categorization of Arabic dialogue processing. In the research, attempts were made to define dialogue processing. It concentrates on classifying words that are used in dialogue. There are various types of dialogue processing, including hello, farewell, thank you, confirm, and apologies. The words are used in the study without context. The proposed approach recovers the properties of function words by replacing collocations with standard number tokens and each substantive keyword with a numerical approximation token. With the use of the linear support vector machine (SVM) technique, the classification method for this study was obtained. The act is classified using the linear SVM technique, and the anticipated accuracy is evaluated against that of alternative algorithms. This study encompasses Arabic dialogue acts corpora, annotation schema, and classification problems. It describes the outcomes of contemporary approaches to classifying Arabic dialogue acts. A custom database in the domains of banks, chat, and airline tickets is used in the research to assess the effectiveness of the suggested solutions. The linear SVM approach produced the best results.

**Keywords**—Dialogue processing; Act; Arabic language; linear support vector machine; without cue

## I. INTRODUCTION

Arabic is recognized as a challenging dialect by non-native speakers of the language in the field of automated language processing. The recognized dialogue acts provide accurate information when the user asks questions and remain silent are typically utilized as an input to the Dialogue Manager component to assist the system decide what to do next. Depending on the dialogue system domain, the Dialogue Acts taxonomy varies [1]. To identify speech acts in dialogues, a variety of techniques have been employed, ranging from rule-based approaches to deep learning and machine learning techniques. To overcome the lack of appropriate training data, novel language interpretation methods have been created [2].

The area of artificial intelligence is employed rarely. Several computer researchers and scientists are engaged in the humanities in order to create computer programmes that accurately scan Arabic writings and convert them into digital formulas. This is because Arabic and artificial intelligence are closely related fields of study. With the advent of new

approaches like Intersent, Elmo, wordszvec and Use, AI was impacted by the introduction of some sophisticated methods for presenting the semantic meaning of texts [3]. In text extraction and processing, conventional ML techniques are typically employed. These techniques include Naïve Bayes, Stochastic Gradient Descent (SGD), and SVM with a linear kernel. The STS problem is addressed using these techniques. Utilizing statistical-based, string and character-based and distance-based similarity metrics, several features are extracted and utilized [4]. The several perspectives for a specific speech sound have typically been represented by statistical methods like Gaussian mixture frameworks [5]. These methods can be employed in a variety of tasks since they are in fact rich in mathematical techniques.

The simplest language communication unit, or dialogue acts (DA) are those that convey the intention of the speakers. Automated dialogue act identification is a crucial indicator for many activities, including subject detection, machine translation, summarization, dialogue systems and human conversation interpretation. Annotation and segmentation are the two key subtasks in dialogue act detection [6]. These two phases can be completed separately—segmentation comes first, then annotation is joined in a single stage. In the context for recognizing dialogue acts, the annotation task is crucial. A label describing the user's intention for every segmented statement during the dialogue is assigned. ML techniques have been tested in studies to detect new DAs. Linear methods and vector-based methods are examples of supervised modeling techniques that are often employed [7]. Limited vocabulary limits are utilized to achieve high speech recognition rates. In reality, it is thought that this outcome is adequate for the majority of voice activation devices to be implemented. As a result, both the learning time and error rate are gradually reduced. Additionally, this rate is plainly erratic and is influenced by lexicon and language. Nevertheless, the creation of advanced ASR systems with increased speech recognition rates has emerged as an intriguing topic of interest for all academics working in the field of speech recognition. In fact, a number of categorization and parametric techniques have surfaced to complete the task. Various categorization techniques, including Gaussian mixture model, Decision tree, machine learning via ANN, K-Nearest Neighbor, Dynamic Time Warping (DTW) and HMM, have been utilized in a number of publications for isolated-words identification systems [8]. But these are inefficient and consume time while attaining the results.

In the Speech Recognition area, computer scientists provided particular consideration for creating effective



techniques to recognize spoken words. Because of the considerable advancements made in this area, Automatic Speech Recognition (ASR) technique is widely utilized. Voice recognition technology called ASR has been designed to assist individuals to communicate better with one another. By utilizing these ASR technologies, it is possible to overcome the challenges imposed on the world's various dialects. As a result, a number of strategies have been developed in the ASR field. An effective method for dialogue processing is the Support Vector Machine [9]. Support Vector Machine (SVM) is used as an estimator of posterior probabilities to improve the performance of identification systems as it has a strong predictive capacity and discrimination. They also have a structural risk minimization (SRM) foundation, where the goal is to create a classifier that minimizes a limit on the predicted risk instead of the empirical risk [5]. Segregating classifier of SVM is considered to have a few practical applications. An algorithm of machine learning for binary categorization is SVM categorization [8]. For multiclass assignments in the real life, it is necessary to change the system's decision-making portion.

SVM is initially intended to function in situations when data has only two classes. In other terms, binary classifier issues can be solved using it. The goal of the multiclass SVM issue is to label cases by selecting labels from a limited number of various elements. Reducing the single multiclass issue into many multiple binary classification issues is the conventional method for utilizing SVM to solve this issue. Building one-versus-all classifiers and selecting the class that correctly categorizes the test examples with the largest margin of error is the most widely used strategy in practice [10]. The presented method first transforms the reader's sound waves into Mel-Frequency Cepstrum Coefficients (MFCC) features before generating a features vector matrix. To train the modified SVM-based technique, portions of the retrieved features are utilized. The trained SVM performed really well when tested twice with the other portion of the collected information. The suggested SVM-based technique was examined employing real-world data. When assessing and contrasting the findings, the tests had very good outcomes. Utilizing the identical datasets of gathered waveforms, the outcomes of the presented SVM-based technique were evaluated with those of other methods for comparison study. The developed SVM-based method surpasses competing methods in terms of accuracy.

**Background:** The situation is complicated, nevertheless, by the degrading language used in speeches delivered by professionals. As a result, it becomes vital to create new tools to aid in the comprehension of flawed Arabic discourses. Speech that is illogical will then be automatically corrected [11]. In Arabic, there are four divisions: Model Standard Arabic (MSA), which is used in newspapers, books, broadcasts, documentaries, formal situations, or when the reader or audience is from a different nation of Arabic, and Dialectal Arabic (DA), which is a dialect of everyday experiences, are all still in use today. Ancient Arabic is no longer in use [12]. In formal situations like journalism, presentations, and courtrooms, the MSA is widely trained in schools and colleges. All countries that speak Arabic

acknowledge MSA as their primary language. Additionally, people frequently use their own languages in everyday conversation. These languages are taught in schools, but there is no organized written record of them. Arabic dialects have extremely distinct morphology, vocabulary, phonology, and syntax from MSA. Dialectal Arabic is the language that is organically spoken in daily life (DA). It varies from one country to the next and can be seen within a country [13]. The user's inquiries and silence are often used as input to the Dialogue Manager component to help the system decide what to do next. Recognized dialogue acts give accurate information when the user asks questions and remains silent.

The remainder of this strategy is distributed into the succeeding divisions: Section II presents the pertinent works and provides a comprehensive analysis of them. Section III describes the problem statement. A thorough review of the proposed method and analysis for the prediction of act classification is provided in Section IV. The results of the experiment are given, examined, and in-depth assessed in Section V, along with a comparison. Finally, the paper is concluded.

## II. RELATED WORKS

The paper [14] utilized DL-based method to show how the meaning of the original and suspect reports was identical. By projecting each potential investment onto its neighbors, the word2vec method did in fact identify the pertinent information. Sentence relating to various could then be produced by combining the acquired dimensions. CNN was then employed to gather more environmental data and determine the level of measuring the similarity. Due to the dearth of accessible to the public material, an edited collection was created using the skip gramme approach. It performed better when substituting an original term with one from language skills that was most comparable and belonged to the same grammatical class. In terms of precision (85%) and recall (86.8%), the work improved efficient environmental connection recognition between Arabic materials compared to earlier studies. for measuring semantic similarity and document modelling, CNN model with various statistical regularities was utilized. A word2vec-based Arabic paraphrased corpus was created for the trials due to the dearth of Arabic paraphrased materials that were accessible to the general audience. Every word from the OSAC source corpus was replaced with its vocabulary counterpart that was the most similar and shared the same grammatical class.

The automatic comprehension of Arabic dialogues for the Egyptian dialect at the utterance level is a challenge that addresses in the paper [15] utilized a machine learning approach, called YOSR. The extractor has been tested on a dataset of impromptu conversations and instant messages for the Egyptian dialect. The results from the YOSR classification are really encouraging. These are the first outcomes for comprehension of the Egyptian dialect that have been recorded, as far as aware. As a viewpoint, we intend to enhance YOSR by including standard call-center regarding the service, context-based characteristics, regional phrase modifications, and morphological information such the initial verb form and Lemma. In order to enhance the categorization

outcomes, also plan to expand the "JANA" dataset for training. However, the efficiency of the method is comparatively slow down the process.

Modern Arabic natural language processing, or ANLP solutions are being developed using machine learning techniques. ML algorithms are commonly employed in NLP because to their high level of accuracy regardless of how strong the input signal is and how simple they are to utilize. On the other hand, the method used in ML-based ANLP implementations entails a number of phases. This evaluation clarifies the concept in detail, illustrates how ML techniques were used to develop these tools, and identifies well-liked ANLP techniques. The paper [16] covers the importance and specifications of ANLP as well as the characteristics and challenges of the Arabic language. Arabic sentiment expressions in tweets can be difficult to distinguish in sentiments categorization software. The complexity of the Arabic script and the unorganized nature of Twitter usage may both contribute to this problem; however, Textual data usability evaluation still needs improvement.

The paper [17] utilized DNN framework was used for sentiment classification, text translation, and text analysis. To text categorization are assessed using a sequence-to-sequence encoder-decoder paradigm, which would consist of NN trained concurrently on outputs and inputs. DNN make use of huge datasets to enhance their accuracy. These linkages are supported by the method, which by identifying textual focal areas, may also be able to manage extensive texts more successfully. This work has re-implementing the basic text useful when analyzing and adapting the sequence-to-sequence structure to Arabic because Arabs have never seen the usage of this approach for text summarizing. About 300,000 elements make up the data set, and each one comprises the title that corresponds with the corresponding article information. After applying standard summarizing techniques on the prior data set, the results are contrasted using the ROUGE index. However, the volume of the AHS information created and raised to the Gig word collection, but it is not much larger than the CNN/Daily Mail database. Additionally, it might make use of other strategies that would be advantageous in the sequential design, such as See et AI demonstrations of a comprehensive approach that uses a various performance vector. The perfect place to concentrate on would be trying to infer new designs that are useful with this vocabulary knowledge in multiple languages, considering that Arabic is a particular textual phrase read from right to left. The research does not take the possibility of increasing the data set into account.

In [18], used an Arabic Alphabet Sign Language Recognition System that is vision-based. Four distinct stages make up the process. The technique can be used with three possible database types: data with hands bound and an alongside a dark tiled wood, information with bare hands and a white base, and data with face buried in a darker-collared glove. By using one of the offered methods in AArSLRS, a hand is first removed from the image and separated from the background before the hand features are eliminated using the search technique that was employed to eliminate them. For the classification of the 28-letter Arabic alphabet using 9240

pictures in this work used methods of supervised learning. To focused on categorizing the 14 special characters and using Quranic sign language to sign the first verses of the Qur'an. It has been proven that picking a resource allocation problem is the optimum choice and that the neighborhood k value of the K-mean clustering algorithm affects rating dependability. However, the method needs more time for the processing phase when compared with other methods.

For an experimental tests of the connection between traditional in the paper [19] employed SR corpus compensation approaches (feature vectors, data selection, sexual identity acoustic frameworks, and dialect-dependent/register-dependent variance across Arabic ASR samples). The first interaction examined in the paper was that between intermittent syntax variance and auditory tracking performance. By removing speakers with inadequate teaching data and switching to grapheme-based acoustic models instead of phone-based ones, discrete specific language difference can be rewarded for. The latter method also helps to make up for poor capture performance, which is further made up for by removing delta-delta acoustic features. Together, the three methods decrease Word Error Rate by 3.24 percentage points to 5.35 percentage. Alteration in the perfectly alright acoustic speech sounds from each phoneme in the phrase is the second feature of regional and registration diversity to be taken into account. Building way of predicting and dialect-specific algorithms contributes to significant reductions in WER since empirical findings show that sexual identity and language are the primary contributors to diversity in speech. Cross-dialect investigations are carried out to gauge how far different Arabic dialects are in terms of the acoustic variations between phone models needed for each of them.

Natural Language Understanding has been vastly enhanced by DL methodologies such as word representations and DNN architectures. This [20] work provides a method for Arabic home automation using DL approaches for text categorization and text categorization recognition. In order to achieve this, to provide an NLU component that can be additionally connected with ASR, a dialogue administrator, and a natural language generator component to develop a fully functional making the change. The procedure of gathering and categorizing the data, constructing the purpose classification and concept extraction frameworks, and eventually the evaluation of these techniques against benchmark datasets are all included in the study. The benchmark results showed that the LSTM effectiveness, with an F-Score of 92.0, was marginally superior to the CNN achievement for both intent classification techniques. The paper employed a hybrid representation of word representations and personality language models, which is then fed to a Bidirectional LSTM network, to retrieve the user objectives and purposes from the information. A high F-Score of 94.0 for the BiLSTM with the Char Embedding's test suggests that its efficiency is extremely comparable to the most recent English Named Entity Recognition benchmarks. However, to create an undertaking dialogue system, the Natural Language Understanding module can be combined with automatic speech recognition and natural language generation modules.

### III. PROBLEM STATEMENT

This issue may be exacerbated by the intricacy of the Arabic script and the disorganised Twitter usage. The evaluation of the usability of textual data still needs work, though. The potential for expanding the data set is not considered in the current research. When compared to alternative ways, the processing step of the current method takes more time. To determine how diverse Arabic dialects are from one another in terms of the acoustic differences between phone models required for each of them, cross-dialect examinations are conducted. However, the Natural Language Understanding module can be integrated with the automatic speech recognition and natural language creation modules to form a complex dialogue system. To develop a classification model for sentential act prediction, the existing technique used a selected group of features extracted from annotated utterances and applied in "YOSR", an SVM technique based on ML. SVM designs ought to be used as a result for improved processing.

### IV. METHODOLOGY

The suggested approach is based on Arabic conversation processing and employs a machine learning algorithm to classify Acts. The contact among the operator and clients in various industries is the subject of the study. The data was gathered from the Jana Corpus in three distinct industries, including banking, travel, and chat. The research's major objective is to establish a connection with the speaker and operator and help them comprehend the dialogue, whether it's a request, an obligation, or something else entirely. Previous

studies have used a variety of algorithms, including deep learning, gradient boosting, convolution neural networks, and natural language processing, but they have not been able to achieve higher accuracy rates. As a result, the research focused on interpreting Arabic dialogue without cues and makes use of chatbot technology and machine learning algorithms to increase productivity. The research is conducted in the stages listed below, and Fig. 1 shows the flow diagram based on the suggested method.

#### A. Dataset Collection

The data collection used in this study was compiled from a variety of sources, including chat, banks, and flights. The vast amount of material on the website is useful for any chatbot application. The CAMEL tool module for Python was used to parse Arabic content and retrieve data. To address a variety of issues, researchers present CAMEL Tools, a user-friendly Python toolkit for Arabic and Arabic dialect pre-processing, morphological modelling, dialect recognition, text categorization recognition, and sentiment analysis. The application programming interfaces (APIs) and command-line interfaces (CLIs) offered by CAMEL Tools cover these functions. This also processes Arabic dialogue using data from chat, planes, and banks. The bot can practise on a dataset, which is a collection of different input expressions and outputs. This effort will allow Arabic programmers to use more dialects of Arabic. The gathered data were split into training and testing samples, with training data making up 70% of the samples and testing data making up 30% of the samples. The sample's division for the testing and training processes is shown in Table I and Fig. 2.

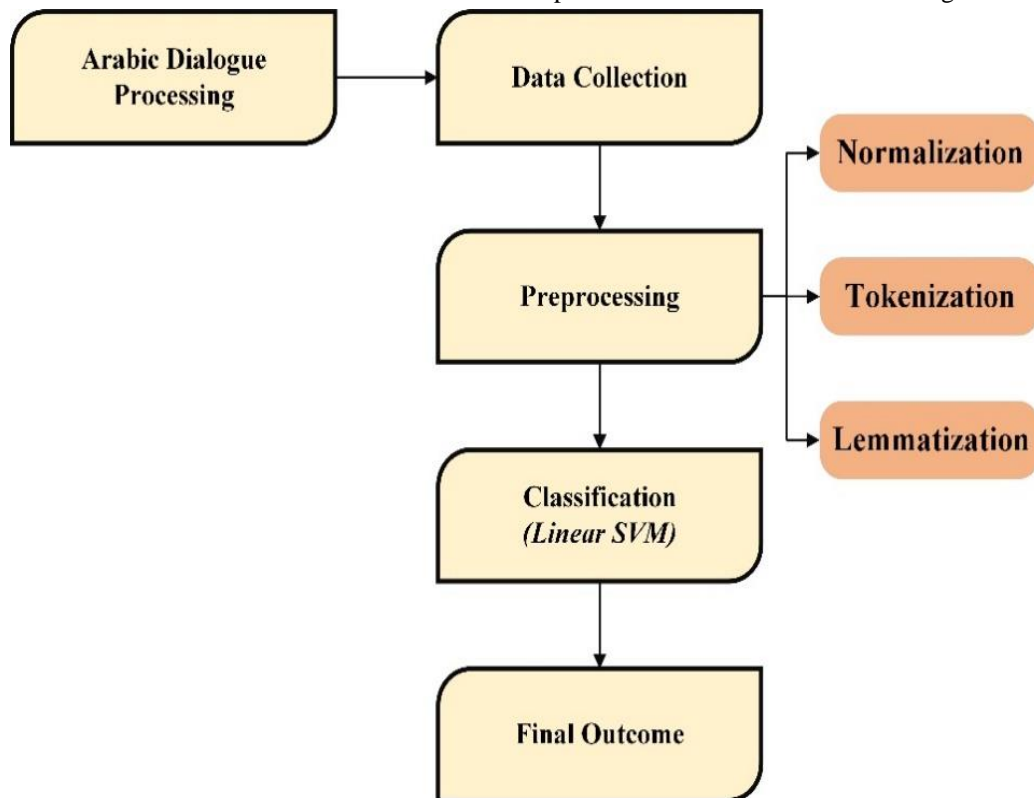


Fig. 1. Proposed method.

Out[31]:

	TID	UID	Text	Speaker	PrimaryClass	SecondaryClass	Act
0	D01T01	D01T01U01	مساء الخير	operator	response	socialobligation	greeting
1	D01T01	D01T01U02	الاهلي فون	operator	other	socialobligation	selfintroduce
2	D01T01	D01T01U03	احمد سامي	operator	other	socialobligation	selfintroduce
3	D01T02	D01T02U01	مساء النور	customer	response	socialobligation	greeting
4	D01T03	D01T03U01	أتشرف بالاسم	operator	request	question	otherquestion
...	...	...	...	...	...	...	...
1133	D56T18	D56T18U01	شكرا ليك اخي سعود	customer	response	argumentation	agree
1134	D56T19	D56T19U01	...أي إستفسارات أو خدمات أخرى اقدر اساعدك فيها ال	operator	request	question	confirmquestion
1135	D56T20	D56T20U01	شكرا	customer	response	argumentation	disagree
1136	D56T21	D56T21U01	- شكرا لتواصلك بخدمه الدعم المياشر	operator	response	socialobligation	thanking
1137	D56T21	D56T21U02	زين عالم جميل	operator	other	socialobligation	selfintroduce

4725 rows × 7 columns

Fig. 2. Input data.

TABLE I. DATA SAMPLE

Data from the JANA Corpus involved in the study	
Transcribed Dialogues	
Telephone Calls	52
Instant Messaging Conversations	30
Total number of words	20311

### B. Preprocessing

Preprocessing's function is to add morphological and other additional context to the lemmatizes (and other following elements') input in order to simplify the lemmatization process. The preprocessing performed by the lemmatization technique is quick and light in comparison to cutting-edge tools. Part-of-speech tagging, performed as a quick machine-learning-based sequence labeler, is used to first enhance the tokenized input text. A very basic word segmented component is used to progressively segment it, reducing the complexity and ambiguity of the words. The procedure's simplicity, which is described in full, can be attributed to the POS tagger's output of a wealth of morphological features, which makes word segmentation essentially easy [21].

In order to make the lemmatization process simpler, preprocessing's role is to add morphology and other contextual info to the lemmatizes (and other following sections') input. The algorithm performs tokenization, function word elimination, and lemmatization for each article in the database. There are numerous Arabic dialects. The analysis of dialect and vernacular terminology follows. Assessments are often written by researchers in their original tongues. Depending on the dialect, they use different terms to convey the same meaning. The solution to this problem was to create and use a dictionary of dialectical terms and their equivalents in standard Arabic [22].

### C. Morphology Analyzer

Morphological analysis is the first stage of text preparation. Breaking down morphologically complicated words into their individual morphemes is known as morphological analysis, also known as analysis of structures (word meaning parts). The morphological analysis of these phrases is then carried out using the Alkhalil analyzer. As a consequence, researchers are provided with all potential lemmas and associated morpho-syntactic information for each individual word in the text that has been analysed out of contextual [23].

### D. Normalization

Experts in Arabic dialogue processing typically employ orthographic normalization as a fundamental technique with the goal of reducing data noise. This is true whether the intention is to produce parallel texts for machine learning, information retrieval data, or computational linguistics text. Examples of normalizing include Tatweel elimination (an effort to eliminate the Tatweel sign), Diacritic removal, and alphabet normalization (variant forms to one form conversion) [24]. The tokenization procedure should become more confusing at this point even if these normalizations would help us in the search or discovering phase. As an illustration, if Ta-Marbuta (ة) and Ha (ه) are normalized, the latter could be tokenized as a pronoun. Because of this, normalizations are treated as a stepping stone for querying, finding, and other operations in the research.

### E. Tokenization

Reducing the raw text into smaller, more manageable chunks is tokenization. It is not an easy problem to solve since tokenization is "closely related to the morphology". This is especially true for dialects like Arabic, which have rich yet complex morphologies. Flowing text is divided into tokens by tokenizers so that they can be processed further by

morphological transducers or POS taggers. The tokenizer recognizes numbers, word boundaries, clitics, multi-word phrases, and abbreviations. By doing this, the original text is divided into tokens, or words and sentences. These tokens support simulation techniques for various processes or contextual comprehension. Tokenization makes it easier to understand the text's significance by looking at the word order. Stop words are typically eliminated in the initial phases of tokenization even though there isn't however a defined list of Arabic stop words [25]. There could be several degrees of Arabic tokenization built, depending on how complicated the linguistic study involved is. Researchers worked with the Arabic grammar to develop three distinct strategies, or frameworks, for Arabic tokenization. In terms of durability, adherence to the modular concept, and ability to remove unnecessary ambiguity, these systems differ greatly from one another. The tokenizer uses white space and punctuation to distinguish among major tokens. However, when defining sub-tokens, the tokenizer needs more morphological information. This information can be provided either deterministically or in-deterministically by a morphology converter or a token forecaster. In the end, both main tokens and sub tokens are identified by the same token borders, which are denoted by the symbol "@" throughout the entire article. It is believed that separating the text into primary tokens using capitalization and white spaces is a straightforward strategy [26].

1) *Utterance feature*: The smallest speech unit in spoken language analysis is termed an utterance. It is a continuous chunk of speech with a distinct pause at the start and the finish. Oral languages typically, but not always, have silence as their boundary. Only their depictions can be found in written language; utterances are absent. The following Table II shows the extraction of utterance.

The utterances' meta data could be used to help with dialogue act classification. Additionally, the classification procedure could benefit from knowing what happened before the current utterance. Researchers made use of:

*Utterance Speaking Style*: The speakers or listeners of the current utterance may be able to discern the act of speech based on their speaking style. The act of "Service-Question," for instance, relates to the customer because they contacted customer service to ask about just one service offered, whereas the acts of "Other-Question" and "Choice-Question" relate to the control system since the operator chooses one of the services offered or asks the client's identity.

*Previous Utterance Act*: The classification may anticipate the act of the present utterance by understanding the sequencing of previous statements in the dialogue. For example, the act "Confirm Question" almost immediately follows the acts "Agree" and "Disagree".

### F. Lemmatization

The main element of our lemmatization strategy is a classifier that uses machine learning. It accepts  $m$ -word segments as input along with their matching POS tag, taking context (utterances and labels) into consideration. The inherent ambiguity of Arabic words without diacritics, whose meanings are often inferred by both individuals and computers from context, justifies the learning-based method. The introduction of new cases necessitates retraining the analyzers. Another disadvantage is that classifications, like the one from OpenNLP that the researcher utilized, frequently settle on an output variable that may or may not be accurate. In this case, additional NLP processing methods from the entire collection of probable lemmas may be able to yield reliable results. To solve these situations, investigators supplement the learning-based lemmatize with a dictionary-based lemmatize.

Fig. 3 illustrates the results of the lemmatization. The analyzer is first used by researchers to look at the words in the database's vowelized version. Once this has been done, researchers only save lemmas whose given lexical tags (clitics, root, and stem) coincide with the word's in-document clitics, root, and stem. The analyzer has located and examined the corpus's words. One potential lemma of the phrases was also discovered during this initial step's study into prospective lemmas. Following the identification of the likely lemmas for each word in the corpus, researchers asked two linguists to select the appropriate lemma from the available options. If the correct lemma is not given among the potential lemmas, the speaker provides the proper lemma to the phrase.

After being lemmatized, the text is again processed to provide the class feature. The class feature obtained in this process is split into final features and final labels, as shown in Fig. 4 and 5. Out from clean text, the first word is acquired, the first verb is filtered and the class feature is extracted, and eventually the class feature is again filtered to obtain the last feature and final labels, as shown in Fig. 4 and 5. As part of this procedure, the provided data were divided into training and testing phases, which were used to classify the data using machine learning.

TABLE II. UTTERANCE FEATURE

Data	
Conversation Turns	3001
Average Conversation/turn	6.7
Utterances	4725
Average words/Utterance	4.3

Out[38]:

	TID	UID	Text	Speaker	PrimaryClass	SecondaryClass	Act	Clean Text	First Verb	Utterance Length
0	D01T01	D01T01U01	مساء الخير	operator	response	socialobligation	greeting	مساء خير	الخير	2
1	D01T01	D01T01U02	الاهلي فون	operator	other	socialobligation	selfintroduce	اهلي فون	فون	2
2	D01T01	D01T01U03	احمد سامي	operator	other	socialobligation	selfintroduce	احمد سام	سامي	2
3	D01T02	D01T02U01	مساء النور	customer	response	socialobligation	greeting	مساء نور	النور	2
4	D01T03	D01T03U01	أتشرف بالاسم	operator	request	question	otherquestion	تشرف اسم	بالاسم	2
...	...	...	...	...	...	...	...	...	...	...
95	D03T08	D03T08U01	انا لي شهادات استثمار مجموعة ب في فرع شريف	customer	response	argumentation	inform	ان لي شهادة استثمار مجموعة ب في فرع شريف	شريف	9
96	D03T09	D03T09U01	تمام	operator	response	argumentation	agree	تمام	تمام	1
97	D03T10	D03T10U01	البنك الاهلي	customer	response	argumentation	inform	البنك اهلي	الاهلي	2
98	D03T10	D03T10U02	ممکن احولها الي الشهادة البلاطينية	customer	request	question	servicequestion	ممکن احول الي شهادة البلاطينيه	البلاطينيه	5
99	D03T11	D03T11U01	تمام	operator	response	argumentation	agree	تمام	تمام	1

100 rows × 10 columns

Fig. 3. Text obtained after lemmatization process.

TID	UID	Text	Speaker	PrimaryClass	SecondaryClass	Act	Clean Text	First Verb	Utterance Length	Class Features
0	D01T01	D01T01U01	مساء الخير	operator	response	socialobligation	greeting	الخير	2	Start Start Start
1	D01T01	D01T01U02	الاهلي فون	operator	other	socialobligation	selfintroduce	فون	2	greeting
2	D01T01	D01T01U03	احمد سامي	operator	other	socialobligation	selfintroduce	سامي	2	selfintroduce
3	D01T02	D01T02U01	مساء النور	customer	response	socialobligation	greeting	النور	2	selfintroduce
4	D01T03	D01T03U01	أتشرف بالاسم	operator	request	question	otherquestion	بالاسم	2	greeting
...	...	...	...	...	...	...	...	...	...	...
1133	D56T18	D56T18U01	شكرا ليك اخي سعود	customer	response	argumentation	agree	سعود	4	confirmquestion
1134	D56T19	D56T19U01	أي إستفسارات أو خدمات أخرى ...أقدر اساعدك فيها ال	operator	request	question	confirmquestion	؟	11	agree
1135	D56T20	D56T20U01	شكرا	customer	response	argumentation	disagree	شكرا	1	confirmquestion
1136	D56T21	D56T21U01	شكرا لتواصلك بخدمه الدعم - المباشر	operator	response	socialobligation	thanking	-	6	disagree
1137	D56T21	D56T21U02	زين عالم جميل	operator	other	socialobligation	selfintroduce	جميل	3	thanking

Fig. 4. Class feature.

TID	UID	Text	Speaker	PrimaryClass	SecondaryClass	Act	Clean Text	First Verb	Utterance Length	Class Features	Final Features	Final Labels
0	D01T01	D01T01U01	مساء الخير	operator	response socialobligation	greeting	مساء خير	الخير	2	Start Start Start	operator Start Start Start	[response, socialobligation, greeting]
1	D01T01	D01T01U02	الاهلي فون	operator	other socialobligation	selfintroduce	اهلي فون	فون	2	greeting	operator greeting	[other, socialobligation, selfintroduce]
2	D01T01	D01T01U03	احمد سامي	operator	other socialobligation	selfintroduce	احمد سام	سامي	2	selfintroduce	operator selfintroduce	[other, socialobligation, selfintroduce]
3	D01T02	D01T02U01	مساء نور	customer	response socialobligation	greeting	مساء نور	النور	2	selfintroduce	customer selfintroduce	[response, socialobligation, greeting]
4	D01T03	D01T03U01	أتشرف بالاسم	operator	request question	otherquestion	تشرف اسم	بالاسم	2	greeting	operator greeting	[request, question, otherquestion]
...	...	...	...	...	...	...	...	...	...	...	...	...
1133	D56T18	D56T18U01	شكرا ليك اخي سعود	customer	response argumentation	agree	شكر لي اخ سعود	سعود	4	confirmquestion	customer confirmquestion	[response, argumentation, agree]
1134	D56T19	D56T19U01	أي استفسارات أو خدمات أخرى اقدر اساعدك فيها ال	operator	request question	confirmquestion	اي استفسار او خدمه اخر اقدر ساعد في الله عافي؟	؟	11	agree	operator confirmquestion	[request, question, confirmquestion]
1135	D56T20	D56T20U01	شكرا	customer	response argumentation	disagree	شكر	شكرا	1	confirmquestion	customer confirmquestion	[response, argumentation, disagree]
1136	D56T21	D56T21U01	شكرا لتواصلك بخدمه الدعم - المياشر	operator	response socialobligation	thanking	شكر تواصل خدمه دعمه -	-	6	disagree	operator disagree	[response, socialobligation, thanking]
1137	D56T21	D56T21U02	زين عالم جميل	operator	other socialobligation	selfintroduce	زين عالم جميل	جميل	3	thanking	operator thanking	[other, socialobligation, selfintroduce]

4725 rows × 13 columns

Activate Windows

Fig. 5. Final feature and label text.

### G. Multilayer outcome and Three Layer Hierarchical Classifications

A kind of machine learning called multi-output classification predicts many outcomes at once. When generating any prediction in multi-output classification, the system would generate multiple or more outcomes. In other classes, the system usually forecasts just one result. Three models—one for each layer—are primarily used in the construction of the suggested technique. As a result, researchers believe that proposed method of classification is quicker than existing classification and may be a more effective dialogue act classification model in real-time systems. The second distinction is that, while "this approach" does not change with the number of dialogue acts or classes, the variety of models does change when binary categorization is utilized, as in the researchers' strategy [27].

1) *Three layer hierarchical*: Traditional classification requires the construction of 26 categories, one for each dialogue act, as opposed to hierarchical structure categorization, which is separate. Every word stated in a discussion is communicated to succeeding tiers to determine the next course of action, which is the fundamental tenet of a hierarchical organization. The recommended method classifies dialogue utterances using a three-layer hierarchical framework. In the first layer, the utterance had already been assigned to one of the three fundamental categories—request, answer, or other. The speech was categorized into various discourse actions in the second layer based on its primary

category, which was established or categorized in the first layer. Question, turn management, social responsibility, social courtesy, argumentation, response, and discourse structure are some of them. Fig. 6 illustrates the hierarchical classification that may be made using the multi-class classification technique.

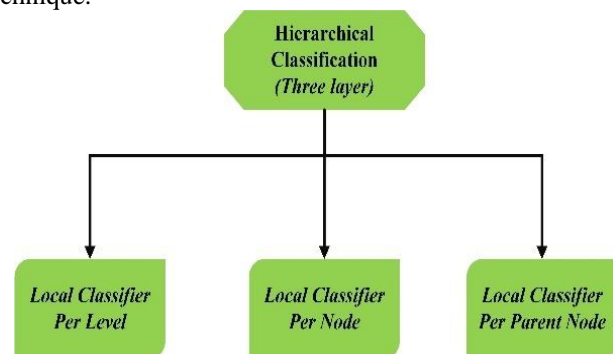


Fig. 6. Hierarchical classification.

Local Classifier per Level:

- For each level of the classification, there is a single multi-class classifier in this system.
- It is a trim, efficient categorization device.

Local Classifier per Node:

- Each node should have a binary classifier.

- Multi-labeling comes easy to this strategy.

Local Classifier per Parent Node:

- Every parent node receiving a single multi-class classifier.
- This is the most logical method.
- This strategy is far more frugal than the prior one.

#### H. Classification

The categorization process is conducted out following the lemmatization technique. In earlier studies, the majority of researchers classify the text using NLP, deep learning, SVM classifiers, gradient boosting, etc. Linear Support Vector Machine is utilized in this study to categorize the provided data. The given dataset was divided into test and training data, and the classification method was applied to the data.

1) *Linear support vector machine*: One of the most popular and well-known machine learning methods is called a support vector machine (SVM). SVMs are supervised learning models that are further used to evaluate and classify text data. SVM is also frequently used for regression evaluation and text classification. Research introduced and applied SVM in text classification and categorization in 1998. The basic purpose of the SVM training method is to create a model that classifies fresh documents into a number of predetermined groups. SVMs could also be employed as non-linear and linear classifiers. A relatively recent class of machine learning algorithms called Support Vector Machines (SVM) was first developed. Based on the data mining and machine learning theory's structure risk minimization concept, SVM seeks a decision surface to divide the training data points into two classifications and bases its judgments on the training examples that are chosen as the only useful components in the training dataset. The SVM Classifier is used for the classification of the input. The Stages convoluted in the procedure are:

- Stage 1: The necessary libraries are added.
- Stage 2: Including the necessary dataset.
- Stage 3: Data pre-processing is carried out.
- Stage 4: Preparing the training and test datasets
- Stage 5: The dataset is encoded to distinguish between labels, and values of 0 or 1 are assigned.
- Stage 6: involves converting text data to polygons. (The user may choose to utilize Term Frequency-Inverse Document Frequency TF-IDF)
- Step 7 involves utilizing SVM to accomplish machine learning.

According to the definition of natural language processing, this describes the conversion of any raw input into highly comprehending. Fig. 7 depicts the pre-processing stages that are involved.

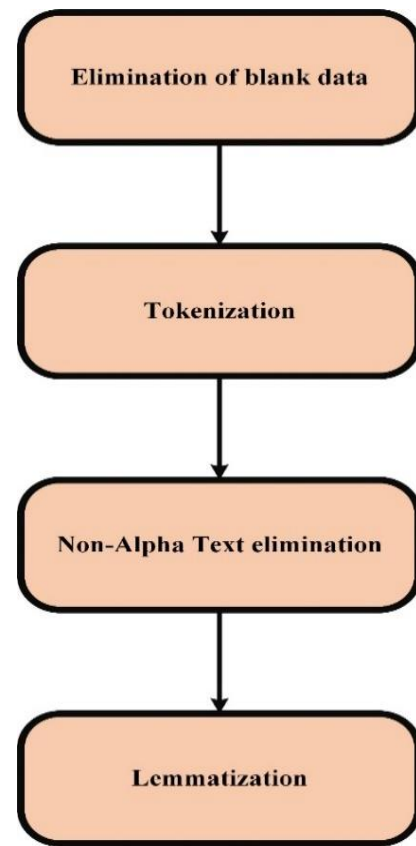


Fig. 7. Preprocessing stages.

#### V. RESULTS AND DISCUSSIONS

Three datasets—the Banks dataset, the Flights dataset, the chat dataset is presented in the experimental findings. SVM, Nave Bayesian, and logistic regression approaches are used independently for each dataset. Moreover, the performance evaluation of different dataset for both multioutput classification and 3-layer hierarchical structure is also presented. Multioutput and local hierarchical classification apply to classifier 3-layer or 3 level structure (primary Categories, secondary Categories and dialogue act ) Primary Categories is (Request, Response, and Other) secondary Categories is (Question, Turn Management, Social Obligation, Social Courtesy, Argumentation, Answer, and Dialogue Structure. First classifier primary categorise next classifier second categories depends on the first classification and last act classification depends on first and second classification.

##### A. Performance Evaluation

Accuracy: Accuracy processes how accurately the system model operates. Usually, it is the proportion of correctly predicted observations to all observations. Accuracy is uttered in eq. (1),

$$A = \frac{\text{Total number of Dialogue act}}{\text{No of dialogue act classified} + \text{No of correct dialogue act}} \quad (1)$$

Precision: Precision is estimated as the number of correct positive estimates of the correct text. It is the fraction of precise identification of the text that is computed utilizing eq. (2),



$$P = \frac{\text{No of correctly classified dialogue act}}{\text{number of dialogue act classified}} \quad (2)$$

Recall: The recall is defined as the ratio of the entire true positives and false negatives to the right positive forecasting accuracy as shown in eqn. (3).

$$R = \frac{\text{No of correctly classified dialogue act}}{\text{number of correct dialogue act}} \quad (3)$$

F1-Score: The F1-score measurement combines precision and recall. Precision and recall are used to calculate the F1-score measure that is symbolized in eqn. (4).

$$F1 - score = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

In the following research Arabic dialogue processing is carried out in machine learning algorithm and the accuracy is obtained and compared with the existing method.

### B. Logistic Regression

The following Table III displays the logistic regression's performance measure for the three parameters of recall, precision, and F1-score for various interactions, including Confirm inquiry, Correct, Disagree, Greet, Inform, misunderstanding sign, Pause, Self-introduce, Service response, Service question, Suggest, accepting request, Thanking, Other answer, other question Warning: Turn assignment.

TABLE III. ACT CLASSIFICATION AND PERFORMANCE MEASURE OF LOGISTIC REGRESSION

Act	Precision	Recall	F1-Score
Agree	0.93	0.80	0.86
apology	0.00	0.00	0.00
Closing	0.82	1.00	0.90
Confirm question	0.34	0.54	0.42
Correct	0.00	0.00	0.00
Disagree	0.63	1.00	0.77
Greeting	0.95	0.90	0.93
Inform	0.15	0.57	0.24
Mis-understanding sign	0.25	1.00	0.40
Offer	0.00	0.00	0.00
Other answer	0.59	0.62	0.60
Other question	0.61	0.86	0.71
pausing	0.38	1.00	0.55
Promise	0.00	0.00	0.00
Repeat	0.00	0.00	0.00
Self-introduce	0.76	0.94	0.84
Service answer	0.92	0.62	0.74
Service question	0.57	0.58	0.58
Suggest	0.00	0.00	0.00
Taking request	1.00	1.00	1.00
Thanking	1.00	1.00	1.00
Turn assign	0.67	0.85	0.75
Warning	0.10	0.75	0.18

### C. Performance Evaluation using SVM Multi-output Classification without Cue

Data are displayed without a cue in Table IV using local support vector machines. Without cues utilizing a support vector machine is explained in Fig. 8. For all datasets, data without cue using support vector machine classification produced better results in terms of recall, precision, accuracy, and f1-score for data without cue. Without employing cue data, a support vector machine was used to achieve an accuracy value of 0.83 and a recall value of 0.85.

A baseline system employing the suggested feature set even without hierarchical characteristics (the major category of the present utterance and the major category of the prior utterance) has been developed in order to assess the efficacy of a Local hierarchical classification in categorizing dialogue acts when used in discussions, flights, and banks. In terms of recall, F measure, precision, and accuracy macro averages, the proposed systems' performances are shown in the table and figure, with multi-output three-layer classification producing the largest result and highest result.

Tables V and VI depict the multioutput classification and local hierarchical structure to interpret data without cue. Performance assessment accuracy findings for bank datasets are higher utilizing Multioutput classification and the Hierarchical structure for supplemented data without a cue. Fig. 9 displays multioutput classification with enhanced data without a prompt. Fig. 10 illustrates data without a cue using a hierarchical structure.

TABLE IV. PERFORMANCE EVALUATION USING SVM

Performance Metrics	SVM Method
Accuracy	0.83
Precision	0.84
Recall	0.85
F1-Score	0.84

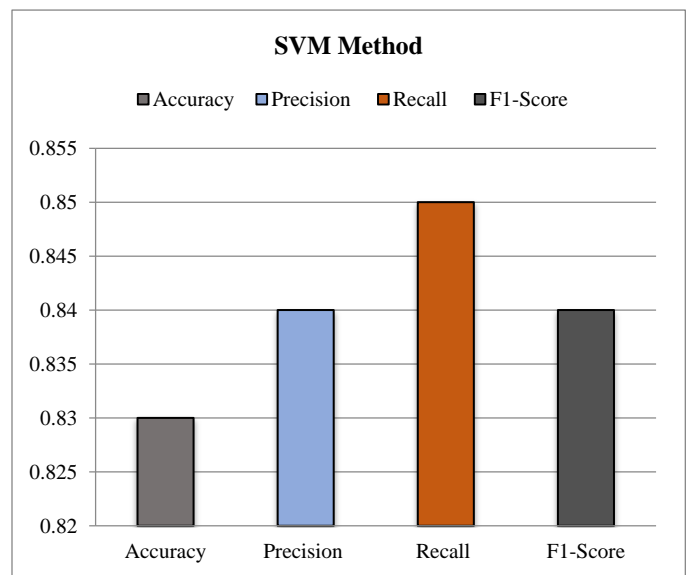


Fig. 8. Performance evaluation using SVM.

TABLE V. MULTI-OUTPUT THREE-LAYER CLASSIFICATION

Multioutput Classification				
	Accuracy	Precision	Recall	F1-score
<b>Banks</b>	0.96	0.95	0.95	0.95
<b>Chats</b>	0.96	0.95	0.94	0.94
<b>Flight</b>	0.95	0.94	0.95	0.94

TABLE VI. THREE-LAYER HIERARCHICAL STRUCTURE

Local hierarchical classification				
	Accuracy	Precision	Recall	F1-score
<b>Banks</b>	0.96	0.95	0.95	0.95
<b>Chats</b>	0.95	0.94	0.94	0.94
<b>Flight</b>	0.95	0.94	0.95	0.94

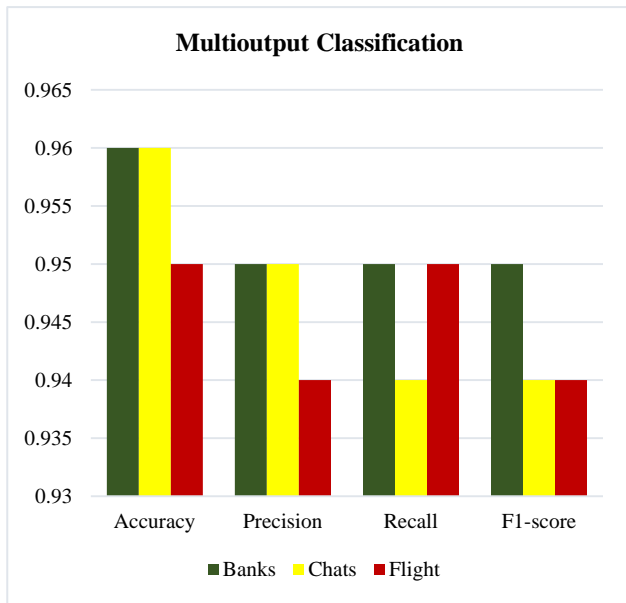


Fig. 9. Multioutput classification.

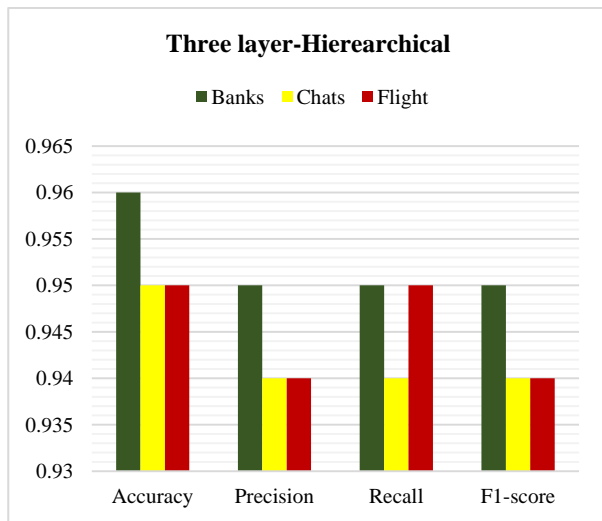


Fig. 10. Local hierarchical classifications.

#### D. Performance Comparison

Compared to the logistic regression approach shown in Table VII, the projected methodology support vector machine achieves a greater level of efficiency. Compared to performance assessed using the logistic regression method, the novel support vector machine produces greater accuracy which shown in Fig. 11. Using the support vector machine model, an accuracy level of 0.83 was achieved in this case. This suggests that a support vector machine is more effective at classifying acts than the logistic regression approach.

TABLE VII. ACCURACY COMPARISON

Method	Accuracy
<b>Logistic Regression</b>	0.72
<b>Linear SV</b>	0.83

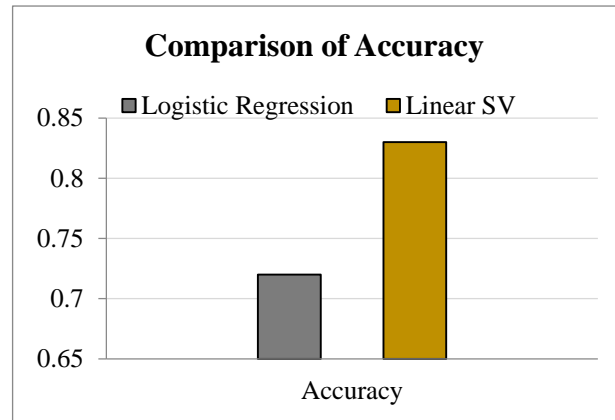


Fig. 11. Accuracy comparison of logistic regression and linear SVM.

#### VI. CONCLUSION

The structure that supports Arabic dialogue processing has been modified for better machine comprehension. The proposed classifier was tested using the Jana corpus of real-world spoken dialogues, and the results are quite positive. Researchers made progress in this work by developing an Arabic Dialogue Processing that will allow users to interact via chat, banking, and flights using textual documents. Additionally, researchers provide the JANA corpus, a multi-genre collection of spontaneous Arabic dialogues that has been tagged at the utterance level in Arabic Dialogues Act Understanding. A technique for classifying utterances according to whether or not they contain function words was created in this work. Among the attributes of this methodology are a database and a set of function words. On the selected dataset, a number of classifiers were tried to see which one would be most useful over the long term. The support vector machine technique is employed in the study to find the customer and operator responses. Better accuracy is attained when comparing the classification system's performance metrics between the logistic regression approach and other methods in the results section. When compared to the other method, the accuracy of the support vector machine method is 0.83. Researchers want to eventually enable clients to make reservations over the phone by connecting our Dialogue Processing with an active booking system, banking, and chat.

To serve as a standard annotated corpus for testing and assessing several Arabic software applications in the future. Finding the appropriate Support Vector Machine parameters can increase accuracy, therefore future research will examine the impact of each parameter. Future study may focus on determining the best approach to integrate various dialogue act classification development techniques in order to enhance the utilization of any technique, even the most basic.

#### ACKNOWLEDGMENT

I'd like to express my heartfelt gratitude to Dr. Md Sah Hj Salam for his insightful and constructive suggestions throughout the planning and development of this research project. His generosity in donating his time has been greatly appreciated. The authors would like to thank Universiti Teknologi Malaysia (UTM) under UTM Encouragement Research grant PY/2020/04246 for the support in the research.

#### REFERENCES

- [1] B. Algotiml, A. Elmadany, and W. Magdy, "Arabic tweet-act: Speech act recognition for Arabic asynchronous conversations," in Proceedings of the Fourth Arabic Natural Language Processing Workshop, 2019, vol. Proceedings of the Fourth Arabic Natural Language Processing Workshop, pp. 183–191. doi: 10.18653/v1/W19-4620.
- [2] Joukhadar, N. Ghneim, and G. Rebdawi, "Impact of Using Bidirectional Encoder Representations from Transformers (BERT) Models for Arabic Dialogue Acts Identification," *Ingénierie Systèmes Inf*, vol. 26, no. 5, pp. 469–475, 2021, doi: <https://doi.org/10.18280/isi.260506>.
- [3] M. M. Khalatia and T. A. H. Al-Romanyb, "Artificial Intelligence Development and Challenges (Arabic Language as a Model)," *Artif. Intell.*, vol. 13, no. 5, 2020.
- [4] M. O. Alhawarat, H. Abdeljaber, and A. Hilal, "Effect of stemming on text similarity for Arabic language at sentence level," *PeerJ Comput. Sci.*, vol. 7, p. e530, 2021, doi: <https://doi.org/10.7717/peerj-cs.530>.
- [5] Zarrouk, Y. Benayed, and F. Gargouri, "Hybrid SVM/HMM model for the recognition of Arabic triphones-based continuous speech," in 10th International Multi-Conferences on Systems, Signals & Devices 2013 (SSD13), Hammamet, Tunisia, Mar. 2013, pp. 1–7. doi: 10.1109/SSD.2013.6564036.
- [6] M. Bashir, A. Hassan, B. Rosman, D. Duma, and M. Ahmed, "Implementation of A Neural Natural Language Understanding Component for Arabic Dialogue Systems," *Procedia Comput. Sci.*, vol. 142, pp. 222–229, 2018, doi: 10.1016/j.procs.2018.10.479.
- [7] S. B. Dbabis, H. Ghorbel, and L. H. Belguith, "Sequential dialogue act recognition for Arabic argumentative debates," 2018, doi: <https://doi.org/10.26342/2018-60-6>.
- [8] W. Mohamed, B. Souha, and C. Adnen, "Speech recognition system based on discrete wave atoms transform partial noisy environment," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, 2019.
- [9] K. M. Nahar, M. Ra'ed, A. Moy'awiah, and M. Malek, "An efficient holy Quran recitation recognizer based on SVM learning model," *Jordanian J. Comput. Inf. Technol. JJCIT*, vol. 6, no. 04, 2020.
- [10] M. Ahed, B. H. Hammo, and M. A. Abushariah, "An Enhanced Twitter Corpus for the Classification of Arabic Speech Acts," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 3, 2020.
- [11] N. Terbeh, M. Maraoui, and M. Zrigui, "Arabic discourse analysis: a naïve algorithm for defective pronunciation correction," *Comput. Syst.*, vol. 23, no. 1, pp. 153–168, 2019, doi: <https://doi.org/10.13053/cys-23-1-2820>.
- [12] K. M. Nahar, O. M. Al-Hazaimh, A. Abu-Ein, and M. A. Al-Betar, "Arabic Dialect Identification Using Different Machine Learning Methods," 2022, doi: <https://doi.org/10.21203/rs.3.rs-1726491/v1>.
- [13] Alsharhan and A. Ramsay, "Investigating the effects of gender, dialect, and training size on the performance of Arabic speech recognition," *Lang. Resour. Eval.*, vol. 54, no. 4, pp. 975–998, Dec. 2020, doi: 10.1007/s10579-020-09505-5.
- [14] Mahmoud and M. Zrigui, "Sentence Embedding and Convolutional Neural Network for Semantic Textual Similarity Detection in Arabic Language," *Arab. J. Sci. Eng.*, vol. 44, no. 11, pp. 9263–9274, Nov. 2019, doi: 10.1007/s13369-019-04039-7.
- [15] A.Elmadany, S. M. Abdou, and M. Gheith, "Towards Understanding Egyptian Arabic Dialogues," *Int. J. Comput. Appl.*, vol. 120, no. 22, pp. 7–12, Jun. 2015, doi: 10.5120/21390-4427.
- [16] S. Larabi Marie-Sainte, N. Alalyani, S. Alotaibi, S. Ghouzali, and I. Abunadi, "Arabic Natural Language Processing and Machine Learning-Based Systems," *IEEE Access*, vol. 7, pp. 7011–7020, 2019, doi: 10.1109/ACCESS.2018.2890076.
- [17] M. Al-Maleh and S. Desouki, "Arabic text summarization using deep learning approach," *J. Big Data*, vol. 7, no. 1, p. 109, Dec. 2020, doi: 10.1186/s40537-020-00386-7.
- [18] Tharwat, A. M. Ahmed, and B. Bouallegue, "Arabic Sign Language Recognition System for Alphabets Using Machine Learning Techniques," *J. Electr. Comput. Eng.*, vol. 2021, pp. 1–17, Oct. 2021, doi: 10.1155/2021/2995851.
- [19] Alsharhan and A. Ramsay, "Investigating the effects of gender, dialect, and training size on the performance of Arabic speech recognition," *Lang. Resour. Eval.*, vol. 54, no. 4, pp. 975–998, Dec. 2020, doi: 10.1007/s10579-020-09505-5.
- [20] M. Bashir, A. Hassan, B. Rosman, D. Duma, and M. Ahmed, "Implementation of A Neural Natural Language Understanding Component for Arabic Dialogue Systems," *Procedia Comput. Sci.*, vol. 142, pp. 222–229, 2018, doi: 10.1016/j.procs.2018.10.479.
- [21] Elmadany, S. M. Abdou, and M. Gheith, "Recent Approaches to Arabic Dialogue Acts Classifications," in *Computer Science & Information Technology ( CS & IT )*, Feb. 2015, pp. 117–129. doi: 10.5121/csit.2015.50412.
- [22] Hmeidi, R. F. Al-Shalabi, A. T. Al-Taani, H. Najadat, and S. A. Al-Hazaimh, "A novel approach to the extraction of roots from Arabic words using bigrams," *J. Am. Soc. Inf. Sci. Technol.*, p. n/a-n/a, 2009, doi: 10.1002/asi.21247.
- [23] S. Larabi Marie-Sainte, N. Alalyani, S. Alotaibi, S. Ghouzali, and I. Abunadi, "Arabic Natural Language Processing and Machine Learning-Based Systems," *IEEE Access*, vol. 7, pp. 7011–7020, 2019, doi: 10.1109/ACCESS.2018.2890076.
- [24] M. M. Kamruzzaman, "Arabic Sign Language Recognition and Generating Arabic Speech Using Convolutional Neural Network," *Wirel. Commun. Mob. Comput.*, vol. 2020, pp. 1–9, May 2020, doi: 10.1155/2020/3685614.
- [25] S. Elons, M. Ahmed, H. Shedid, and M. F. Tolba, "Arabic sign language recognition using leap motion sensor," in 2014 9th International Conference on Computer Engineering & Systems (ICCES), Cairo, Egypt, Dec. 2014, pp. 368–373. doi: 10.1109/ICCES.2014.7030987.
- [26] Joukhadar, N. Ghneim, and G. Rebdawi, "Impact of Using Bidirectional Encoder Representations from Transformers (BERT) Models for Arabic Dialogue Acts Identification," *Ingénierie Systèmes Inf.*, vol. 26, no. 5, pp. 469–475, Oct. 2021, doi: 10.18280/isi.260506.
- [27] Elmadany, S. Abdou, and M. Gheith, "Improving Dialogue Act Classification for Spontaneous Arabic Speech and Instant Messages at Utterance Level," 2018, doi: 10.48550/ARXIV.1806.00522.

# Tree-based Machine Learning and Deep Learning in Predicting Investor Intention to Public Private Partnership

Ahmad Amin<sup>1</sup>, Rahmawaty<sup>2</sup>, Maya Febrianty Lautania<sup>3</sup>, Suraya Masrom<sup>4</sup>, Rahayu Abdul Rahman<sup>5\*</sup>

Faculty of Economics and Business-Universitas Gadjah Mada, Yogyakarta, Indonesia<sup>1</sup>

Faculty of Economics and Business-Universitas Syiah Kuala, Aceh, Indonesia<sup>2,3</sup>

Computing Sciences Studies-College of Computing-Informatics and Media, Universiti Teknologi MARA Perak Branch, Malaysia<sup>4</sup>

Faculty of Accountancy, Universiti Teknologi MARA, Perak Branch, Malaysia<sup>5</sup>

**Abstract**—Public private partnership (PPP) is the government initiate in accelerating public infrastructure development growth. However, the scheme exposes private sector to various risks including political risk which in turn affect financial performance and reporting of participating firms. Given that one of the issues facing the government is the lack of participation from the private sector in such arrangements. Thus, the main objective of this study is to observe the machine learning prediction models on private investor intention in participating the PPP program. Tree-based machine learning and deep learning are two different types of promising algorithms, which proven to be useful in widely domain of prediction problems but never been tested on the concerned problem of this study. Based on real data of investors for Indonesian listed firms, this paper presents the ability of the selected machine learning algorithms by means of different assessments point of view. First assessment is on the algorithms' performances in producing accurate prediction. Second assessment is to identify the variance of PPP attributes in each of the prediction model with the machine learning algorithms. The performance results show that all the prediction models with the machine learning algorithms and the PPP attributes were well-fitted at R squared above 80%. The findings contribute a significant knowledge to various fields of scholars to implement a more in-depth analysis on the machine learning methods and investors' prediction.

**Keywords**—Tree-based machine learning; deep learning; prediction; investor intention; public private partnership

## I. INTRODUCTION

The degree of development of a nation is determined by its capacity to provide for the needs of the general population, including infrastructure like power, internet, trains, roads, ports, and airports. Infrastructure investments may improve livelihoods and well-being since they make people's lives easier and better [1], [2]. However, many countries have several issues and difficulties when constructing public infrastructure. Financial factors, regional factors, demographic concerns, environmental difficulties, and human factors are some of the key causes of infrastructure development delays [3], [4].

In response to the financial constraint, the government has

initiated alternative funding programmes for the construction of public infrastructure and invited the private sector to participate. This special arrangement is known as public-private partnerships (PPP). PPP has developed in a number of developing countries since it was first used in the United Kingdom in the early 1990s, including Chile in Latin America, Colombia, Brazil, Bulgaria, and Slovenia in Eastern Europe, and Indonesia, Malaysia, and South Korea in Asia [5]–[7]. However, the degree to which PPPs are successfully implemented varies from country to country. For example, in Indonesia, despite the surge in PPP projects for infrastructure development, the private sector's participation is still low as the scheme dominated by Indonesian state-owned enterprise. Hence, this study aims to develop machine learning prediction model on private investor intention in participating the PPP program. This study has important contributions by extending prior works on PPP with non-financial factors to be analyzed based on the machine learning prediction models.

The following section provides a brief description on the data set of this study and the empirical steps of the machine learning research. Section III presents the results of the machine learning performances comparisons based on the PPP attributes. Finally, Section IV discusses the conclusions and future research directions.

## II. LITERATURE REVIEW

Prior research that applied machine learning on prediction, classification and detection studies in financial, accounting and finance research highlight the effectiveness and accuracy of such methods to that of traditional statistical methods in problems such as in detection of financial fraud [8], firm performance [9] and finance [10], [11]. Despite the wider use of machine learning in accounting and finance, yet study on machine learning prediction and classification on PPP investment is limited. To date, most of prior studies used machine learning to predict successful PPP projects [12]–[16] that highlighted the return benefits of deploying the intelligent approach for solving various issues of PPP. In [16], Random Forest has been identified as the most outperformed algorithm but the study has observed different aspects of PPP attributes that used in this study. Most of the researchers suggested that more advance research on the machine learning algorithms for

---

This research was funded by Ministry of Higher Education Malaysia under FRGS grant 600-IRMI/FRGS 5/3 (208/2019).

PPP is needed to accelerate the implementation of computational system for facilitating PPP projects and investments.

Prior studies on the determinants of PPP highlight several financial, economic and environmental factors such as general and operational risk [17],[18], return [18]; operational and cost recovery [19], ownership [20], risk of political instability [21]; risk of exogenous uncertainty [22] and environmental, development impacts and sustainability [20]. However, there are limited studies that have been conducted to examine how non-financial factors such as investor trust in the government, influence the decision of investors to finance public infrastructure through PPP arrangements. To provide a deepen current understanding on the private investor intention in participating in PPP programs, this study attempted to discover non-financial factors including government trust, service quality, transparency, and values of similarity as the determinants of PPP. Unlike the prior works, this paper presents the findings of analysis from the implementation of machine learning prediction models.

To date, more than hundreds of machine learning algorithms can be utilized for various domains of problem such as Decision Tree [23], Random Forest [24], Support Vector Machine [24] and Gradient Decision Trees [23]. Decision Tree, Random Forest and Gradient Decision Trees are categorized as tree-based machine learning. The tree-based machine learning is robust to be used for regression and classification problems. Additionally, Deep Learning [25] is very promising algorithm for prediction problems with massive attributes. However, to study its ability on simple attributes as proposed in this study will provide another useful knowledge to researchers from various fields of interest. The findings will be beneficial to the relevant PPP stakeholders for implementing vast and rapid data-driven recommendations and decisions.

### III. METHODS

#### A. Dataset

The machine learning algorithms were tested on dataset that consists of data from 165 top management of Indonesian listed firms. Based on the real collected data, Pearson correlation test was conducted to observe the weights of correlation coefficient each of the PPP attributes as the independent variables (IVs). Fig. 1 presents the PPP attributes correlation coefficient to the dependent variable (DV). The DV in the predictor model is the investor intention to the PPP. The PPP attributes is the investment intention factors namely based on government trust (*TrustOnGov*), perceived of government service quality (*GovServiceQ*), perceived of government transparency (*GovTransparency*), and similarity of values (*ValueSimilarity*).

Two of the PPP attributes present positive strong correlations (above 0.7 correlation coefficient) to the investor intention while the rest of two attributes have moderate correlation coefficient (0.5-0.6). In the machine learning predictive models, one important matter that need to be observed is the contribution of each attribute in providing

knowledge for the machine learning to make prediction. Two questions to be answered in this research are:

RQ1: Are all the four PPP attributes usable to each of the machine learning prediction model?

RQ2: Which of the attributes are mostly important to all the machine learning models?

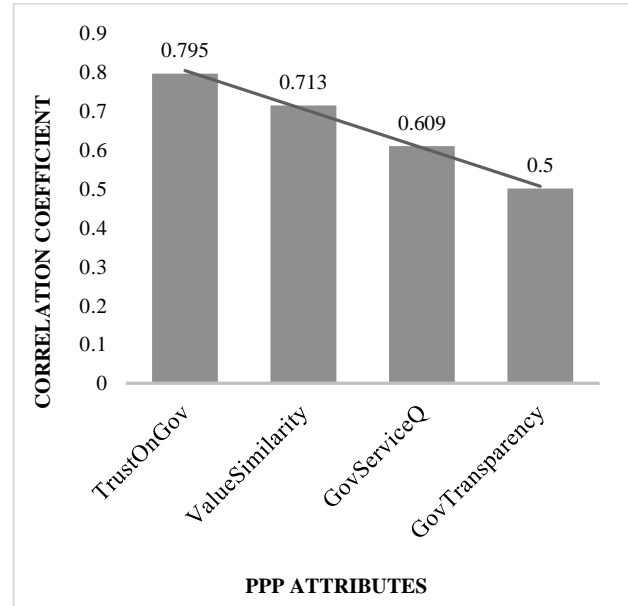


Fig. 1. Weights of correlation each PPP attributes to the DV.

#### B. Machine Learning

Two machine learning from the family of tree-based algorithms namely Random Forest (RF) and Decision Tree (DT) were used in this study. Additionally, acknowledging the wider used of Deep Learning algorithm, it is interesting to observe the algorithm ability in the case of PPP investor intention. Based on preliminary experimental series, the most optimal hyper-parameters of RF and DT is listed in Table I.

The range of number of trees for observing the RF accuracy were 20, 60, 100,140. For each combination, the worst error rate produced by RF was 11.6% while the best error rate was 10.5% with number of tress 140 and maximal depth 4. For DT, the range of maximal depth used in the preliminary testing is between 2 and 25. The most optimal value has been achieved by DT when the maximal depth was 4 at error rate 11.8%, slightly lower from the maximal depth value 2(11.9%). A similar value of error rate at 12.8% has been produced when the maximal depth was individually set as 7, 10,15 or 25. The production of prediction generated by Deep learning is depicted in Fig. 2. The number of layers is 4 that used Rectifier function for the input layers and Linear function as the output layer (layer 4).

TABLE I. THE OPTIMAL HYPER-PARAMETERS

Machine Learning Algorithm	Hyper-parameters	Error rate %
RF	Number of trees=140 Maximal depth=4	10.5
DT	Maximal depth=4	11.8

```

root mean squared log error: 0.106737055
Status of Neuron Layers (predicting InvestmentIntention, regression, gaussian distribution, Quadratic loss, 2,851 weights/biases, 38.3 KB,
Layer Units      Type Dropout      L1      L2 Mean Rate Rate RMS Momentum Mean Weight Weight RMS Mean Bias Bias RMS
1      4      Input  0.00 %
2      50 Rectifier  0 0.000010 0.000000 0.002733 0.001804 0.000000 -0.003671 0.187752 0.496365 0.013304
3      50 Rectifier  0 0.000010 0.000000 0.007516 0.014828 0.000000 -0.001072 0.138856 0.997919 0.013785
4      1      Linear  0.000010 0.000000 0.000333 0.000157 0.000000 0.024648 0.204335 -0.007682 0.000000

Scoring History:
Timestamp      Duration Training Speed      Epochs Iterations      Samples Training RMSE Training Deviance Training MAE Training r2
2022-11-02 10:33:25 0.000 sec      0.000000      0 0.000000      NaN      NaN      NaN      NaN
2022-11-02 10:33:25 0.013 sec      20625 obs/sec 1.000000      1 165.000000 0.49599      0.24600      0.37078      0.56758
2022-11-02 10:33:25 0.024 sec      19411 obs/sec 2.000000      2 330.000000 0.53033      0.28125      0.40805      0.50562
2022-11-02 10:33:25 0.034 sec      19038 obs/sec 3.000000      3 495.000000 0.45317      0.20537      0.34044      0.63901
2022-11-02 10:33:25 0.045 sec      18857 obs/sec 4.000000      4 660.000000 0.46653      0.21765      0.34810      0.61742
2022-11-02 10:33:25 0.055 sec      19186 obs/sec 5.000000      5 825.000000 0.47776      0.22826      0.36527      0.59877
2022-11-02 10:33:25 0.065 sec      19411 obs/sec 6.000000      6 990.000000 0.44285      0.19612      0.32818      0.65527
2022-11-02 10:33:25 0.076 sec      19250 obs/sec 7.000000      7 1155.000000 0.44362      0.19680      0.32992      0.65407
2022-11-02 10:33:25 0.086 sec      19411 obs/sec 8.000000      8 1320.000000 0.43911      0.19282      0.32633      0.66106
2022-11-02 10:33:25 0.098 sec      19285 obs/sec 9.000000      9 1485.000000 0.44436      0.19745      0.33222      0.65292
2022-11-02 10:33:25 0.112 sec      18333 obs/sec 10.000000     10 1650.000000 0.44296      0.19621      0.33480      0.65511
2022-11-02 10:33:25 0.117 sec      17741 obs/sec 10.000000     10 1650.000000 0.43911      0.19282      0.32633      0.66106

```

Fig. 2. The production of deep learning.

#### IV. RESULTS AND DISCUSSION

Table II presents the performances comparison between the three machine learning algorithms by means of Root Mean Square Error (RMSE), R squared ( $R^2$ ) and the time required to complete (TTC) the processes of training and predicting. RMSE measures the accuracy error of the machine learning models while  $R^2$  is presenting the proportion of IVs variation in the prediction models.

Table II shows that all the predictive models with the PPT attributes have contributed at least 80% of the fitted data from the  $R^2$ . Random Forest provided the best results in line with the results in [16] but on this cases that used different PPP attributes, the  $R^2$  is lower that the results in [16]. Although the average error presented by the Relative Error and RMSE of all models are not very encouraging, the values are within an acceptable range to anticipated that the machine learning can relatively predict the data accurately. All the algorithms are reliable when tested on the PPP dataset as described through the results of small standard deviation ranges.

In term of efficiency, DT has taken the shortest time to complete followed by Deep Learning and RF. Furthermore, it will be useful to look the variance of each PPT attributes in the machine learning predictive models as listed in Table III.

As seen in Table III all the PPT attributes were contributed some degree of correlations in the prediction models of all the machine learning algorithms. Thus, all the PPP attributes used in this study were used by the machine learning algorithms, than answered the RQ1. Deep Learning received bigger weight of contributions from the trust on government and value of similarity. For all algorithms, trust on government has given a major influence the prediction models at weight above 0.5, which answering the RQ2 of this research. The lowest influence from the PPP attributes can be seen from the government service of quality. Furthermore, the following Fig. 3 and Fig. 4 show the Tree models generated by DT and RF, respectively.

TABLE II. THE PERFORMANCES RESULTS

Algorithm	Relative Error (+-std.dev)	RMSE (+-std.dev)	$R^2$ (std.dev) (+-)	TTC (ms)
RF	9.9% (3.8%)	0.487 (0.108)	0.826 (0.094)	553
DT	10.3% (2.5%)	0.5 (0.108)	0.802 (0.114)	61
Deep Learning	10.4% (2.5%)	0.477 (0.098)	0.824 (0.056)	342

TABLE III. WEIGHTS OF CORRELATION BETWEEN PPP ATTRIBUTES IN MACHINE LEARNING

Algorithm	Trust on Government (TrustonGov)	Value of Similarity (ValueSimilarity)	Government Transparency (GovTransparency)	Government Service of Quality (GovServiceQ)
RF	0.674	0.395	0.078	0.053
DT	0.597	0.230	0.055	0.048
Deep learning	0.717	0.662	0.028	0.029

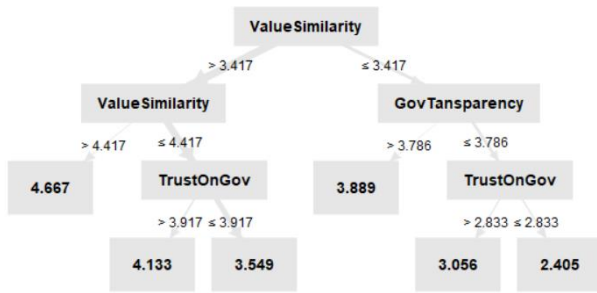


Fig. 3. The decision tree.

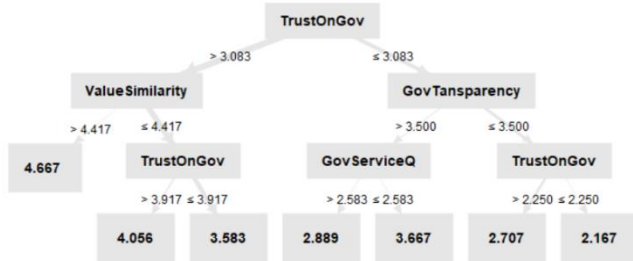


Fig. 4. The random forest tree at 140 numbers of trees.

The depth of the tree models from DT and RF is 4 as set at the optimal maximal depth. The trust on government is the deeper attribute before the leaf of both trees that representing it as the important contribution to the model. Although the weights of similarity of value and government transparency are very low in DT and RF, the two attributes are still included in the tree models. Otherwise, government service of quality is utilized in RF but not in DT.

## V. CONCLUSIONS

This study aims to explore the non-financial issues that have influenced the investor intention in the public-private partnerships (PPP) based on the evidence from Indonesia. Acknowledging the role of machine learning to provide fast data-driven prediction and how complex is the challenge of future data expanding, this study attempted to discover machine learning techniques and has observed the effect of the pre-determined non-financial factors in the successful of PPP including Government Trust, Similarity of Values, Transparency in Government and Service Quality of the Government. Within the scopes of the tested data based on PPP in Indonesia, the attribute of Government Trust presents the most significant factor in the prediction models inside and outside of the machine learning implementations. This study can be further extending in the future by considering more PPP attributes as suggested in [16] that can improve the accuracy of the machine learning models and use different machine learning algorithms than the proposed study in this paper.

## ACKNOWLEDGMENT

This research was supported by the Ministry of Higher Education Malaysia under FRGS grant 600-IRMI/FRGS 5/3 (208/2019). Thank you very much to the Ministry of Higher Education Malaysia and the Universiti Teknologi MARA for the full support.

## REFERENCES

- [1] H. Kociemska, "Public-Private Partnership: Reconciling Mainstream and Islamic Finance in sub-Saharan Africa," *Emerg. Mark. Financ. Trade*, vol. 56, no. 12, pp. 2891–2907, 2020.
- [2] S. Y. In, L. A. S. Casemiro, and J. Kim, "A decision framework for successful private participation in the airport sector," *J. Air Transp. Manag.*, vol. 62, pp. 217–225, 2017.
- [3] I. Walter, "The infrastructure finance challenge," Available SSRN 2841281, 2016.
- [4] S. S. Clark, T. P. Seager, and M. V. Chester, "A capabilities approach to the prioritization of critical infrastructure," *Environ. Syst. Decis.*, vol. 38, no. 3, pp. 339–352, 2018.
- [5] S. Pfisterer, "Public-Private Partnership for Development: Governance Promises and Tensions," in *The Emerald Handbook of Public-Private Partnerships in Developing and Emerging Economies*, Emerald Publishing Limited, 2017.
- [6] R. Mohamad, S. Ismail, and J. M. Said, "Performance objectives of public private partnership implementation in Malaysia: perception of key players," *J. Asia Bus. Stud.*, 2018.
- [7] S. Ismail, M. S. Musawa, and H. Ahmad, "Transparency of public private partnership (PPP): the extent of mandatory information disclosure," *Built Environ. Proj. Asset Manag.*, 2019.
- [8] A. Priyadarshini, S. Mishra, D. P. Mishra, S. R. Salkuti, and R. Mohanty, "Fraudulent credit card transaction detection using soft computing techniques," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 23, no. 3, pp. 1634–1642, Sep. 2021, doi: 10.11591/ijeecs.v23.i3.pp1634-1642.
- [9] P. H. D. Abd Samad, S. Mutalib, and S. Abdul-Rahman, "Analytics of stock market prices based on machine learning algorithms," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 16, no. 2, pp. 1050–1058, 2019.
- [10] M. A. Muslim, Y. Dasril, M. Sam'an, and Y. N. Ifriza, "An improved light gradient boosting machine algorithm based on swarm algorithms for predicting loan default of peer-to-peer lending," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 2, pp. 1002–1011, 2022.
- [11] A. K. Bitto et al., "CryptoAR: scrutinizing the trend and market of cryptocurrency using machine learning approach on time series data," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 3, pp. 1684–1696, 2022.
- [12] S. Ghorbany, S. Yousefi, and E. Noorzai, "Evaluating and optimizing performance of public-private partnership projects using copula Bayesian network," *Eng. Constr. Archit. Manag.*, no. ahead-of-print, 2022.
- [13] Y. Wang and R. L. K. Tiong, "Public-Private Partnership Contract Failure Prediction Using Example-Dependent Cost-Sensitive Models," *J. Manag. Eng.*, vol. 38, no. 1, p. 4021079, 2022.
- [14] Y. Wang, Z. Shao, and R. L. K. Tiong, "Data-driven prediction of contract failure of public-private partnership projects," *J. Constr. Eng. Manag.*, vol. 147, no. 8, p. 4021089, 2021.
- [15] M. Eskandari, M. Taghavifard, I. R. Vanani, and S. S. G. Noori, "An Intelligent Hybrid Model for Determining Public-Private Partnership in Iranian Water and Wastewater Industry Based on Collective Tree Algorithms," *J. Water Wastewater*, vol. 32, no. 1, pp. 69–90, 2021.
- [16] Y. Wang, Z. Shao, and R. L. K. Tiong, "Data-driven prediction of contract failure of public-private partnership projects," *J. Constr. Eng. Manag.*, vol. 147, no. 8, p. 4021089, 2021.
- [17] L. L. Martin, "Making sense of public-private partnerships (P3s)," *J. Public Procure.*, 2016.
- [18] G. Y. Debela, "Critical success factors (CSFs) of public-private partnership (PPP) road projects in Ethiopia," *Int. J. Constr. Manag.*, vol. 22, no. 3, pp. 489–500, 2022.
- [19] F. Calabrò and L. Della Spina, "The public-private partnership for the enhancement of unused public buildings: an experimental model of economic feasibility project," *Sustainability*, vol. 11, no. 20, p. 5662, 2019.
- [20] D. Albalate, G. Bel, and R. R. Geddes, "Recovery risk and labor costs in public-private partnerships: Contractual choice in the US water industry," *Local Gov. Stud.*, vol. 39, no. 3, pp. 332–351, 2013.

- [21] C. B. Casady, K. Eriksson, R. E. Levitt, and W. R. Scott, "(Re) defining public-private partnerships (PPPs) in the new public governance (NPG) paradigm: an institutional maturity perspective," *Public Manag. Rev.*, vol. 22, no. 2, pp. 161–183, 2020.
- [22] C. Oliveira Cruz and J. Miranda Sarmiento, "Theory on PPP renegotiations," in *The renegotiations of public private partnerships in transportation* Springer, 2021, pp. 19–27.
- [23] M. Vaudevan, R. S. Narayanan, S. F. Nakeeb, and Abhishek, "Customer churn analysis using XGBoosted decision trees," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 25, no. 1, pp. 488–495, 2022, doi: 10.11591/ijeecs.v25.i1.pp488-495.
- [24] R. Meenal, P. A. Michael, D. Pamela, and E. Rajasekaran, "Weather prediction using random forest machine learning model," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 22, no. 2, pp. 1208–1215, 2021, doi: 10.11591/ijeecs.v22.i2.pp1208-1215.
- [25] X. Hao, G. Zhang, and S. Ma, "Deep learning," *Int. J. Semant. Comput.*, vol. 10, no. 03, pp. 417–439, 2016.



# VHDL based Design of an Efficient Hearing Aid Filter using an Intelligent Variable-Bandwidth-Filter

Ujjwala S Rawandale<sup>1</sup>, Dr. Sanjay R. Ganorkar<sup>2</sup>, Dr. Mahesh T. Kolte<sup>3</sup>

Dr. Vishwanath Karad MIT World Peace University and Research Scholar at Sinhgad College of Engineering<sup>1</sup>

Electronics and Telecommunication Engineering Department, Sinhgad College of Engineering<sup>2</sup>

Electronics and Telecommunication Engineering Department PCCOE<sup>3</sup>,

Pune, Maharashtra, India - 411038

**Abstract**—Filtering techniques have been elaborated in the HA field to improve signal clarity and enhance the hearing capacity of deaf people. However, public sounds are highly noisy, so filtering those signals is not an easy task. Hence, the present article has aimed to develop a novel Ant Lion based power Noise-Variable Bandwidth Filter (ALPN-VBF) for the HA applications. Here, the proposed optimized power efficient filter has incorporated several functions like de-noising and frequency tuning based on the word features. Here, the signal's noise has been removed with the maximum possible range with the help of High-pass-Filter (HPF) and low-pass filter (LPF). Finally, the developed model is tested with a few audiograms, and the filter parameters have been analyzed and compared with other models. The testing results have proved that the designed filter is better in frequency tuning and signal transmission than the previous approaches by attaining less delay and reduced power consumption rate.

**Keywords**—Hearing aid system; variable bandwidth filter; audiograms; matching error; power consumption; signal filtering

## I. INTRODUCTION

The response variable frequency properties are required in various telecommunications applications and software wireless applications [1]. In addition, the design of bandwidth filter with variable parameters has included adders, coefficient array, multipliers, and data array length [2]. The idea employed here reduces the relative filter bandwidth by functioning it in a lower sample rate [3]. Moreover, this process is balanced by an efficient resembling technique. Here, the down-sampled wave is treated by a fixed bandwidth and length, and the up-sampled return to the initial spectrum [4]. Hence, the combined strategy has decreased the filter's bandwidth [5]. Moreover, this approach is chiefly utilized in bandwidth reduction functions [6]. However, numerous applications may demand a higher bandwidth than the fixed-length filter's performance [7]. However, this could result in distortion of signal response outcome [8]. Besides, the non-zero bandwidth transition is the primary cause of distortion in fixed-length-filter [9]. The filter's length often depends on the bandwidth transition. In addition, to gain the undisturbed signal, reducing the filters distortion has to be minimized up to the desired level [10].

The chief advantage in using VBF is better tuning capacity based on the signaling and noise application. The resonant of frequencies have been tuned accordingly [11]. The basic function of VBF is detailed in Fig. 1. In addition, the complexity in VBF is noise removal [12]. If the data is

realistic, it has taken more resources and time to de-noise the signal; also, the conventional filter has afforded the average outcomes [13]. Several filtering approaches, such as a novel key element based on cloud serve VBF [14], parallel structure filter [15], etc., have been implemented in the past. But, the appropriate filtered output was not gained, so the current research work has aimed to design a novel power-efficient, optimized filter model for HAA application. At last, the performance of the designed filter is compared with other models and has obtained the finest results. The key contributions followed in this research is described as follows:

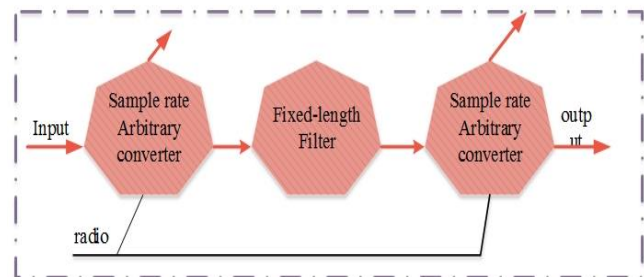


Fig. 1. Function of VBF

- Initially, different public audios were collected from the benchmark site and trained to the system.
- Consequently, a novel ALPN-VBF has been developed with suitable parameters to denoise and tune the signal frequency.
- Here, the possible level of noise signal has been reduced with the use of the HPF and LPF.
- Moreover, the desired signal range has been tuned by matching the maximum signal range in the ant lion fitness.
- Subsequently, the incorporation of the ant lion fitness in frequency tuning has afforded the finest outcome.
- Finally, the designed parameters were measured and compared with conventional paradigms in the relations of power consumption, delay, and matching error.

The current article's research arguments are organized as follows. Section II has detailed the associated works of the different filters in the HA system; the designed novel VBF

filter and function is highlighted in Section III. Moreover, the working performance of the designed Filtering strategy is summarized in Section IV, and the research discussion and achievements are concluded in Section V.

## II. RELATED WORKS

To enhance the tuning range of VBF, Ren et al. [14] have implemented a novel key element based on cloud services for the VBF system. The designed VBF has recorded the tuning rate between 0.3nm to 1.35nm by the experimental process. Also, the measured loss of insertion is 1.39dB. Hence, the tuning function of the VBF has been maximized. Also, it is flexible to change the response rate based on applications. However, it has required more power than the conventional VBF.

The parallel structure filter has been introduced by Swamy and Alex [15] for the HA system for better signal; frequency. Here, the filter is modeled with the reconfigurable parameters that contain several sub-bands to process the different frequency range signals. Moreover, this filter is tested with numerous kinds of Hearing Loss (HL) people data. Hence, the tested people data contains mild to severe HL data; in many cases, the signal has reached better. However, it has consumed more power.

The corrector filters with half bands have been described for the HA system by Niveditha et al. [16] to enhance signal transmission speed. The working of the designed filter is verified with the hardware HA application. The proposed filter has earned a reduced power consumption rate and high signal transmission speed. However, it has lacked in signal clarity.

Ma et al. [17] modeled the reconfigurable filter with multi bands and modularized design structure. In addition, the multiple distortion schemes of subbands have met the frequency range of several different audiograms. The designed filter has a reduced complexity rate than the conventional noise removal filter in HA systems. However, it has required more energy to execute the process.

The notch filter has been applied in the HA system by notch filter Marcum et al. [18] to maximize the hearing capacity of the aged hearing loss people. The key motive of this research is to measure the filtering rate by comparing the conventional filter. Hence, the filtering capacity was identified by using a different set of audio signals. However, it has required more resources to execute.

Several methods were discussed with its limitation and merits for the HA system. Hence, the common demerit in many cases is signal noise. If this problem ends, then the HA system's signal processing has been maximized. So, the current work has focused on implementing the powerful noise filter for HA applications.

## III. PROPOSED METHODOLOGY

The key motive of this HA system application is to improve the communication and hearing capacity of deaf people. Several kinds of HA system has been designed with different efficient filter parameters. But, still, the noise in the processing signal might reduce the signal frequency. So, the present work

has focused design a novel optimized VBF named ALPN-VBF for HA framework. Here, to enable the power noise mechanism, dual filters functions have been utilized that are HPF and LPF. Then to transform the optimal signal range, the maximum signal frequency range is modeled in the Ant lion fitness function. Hence, the optimization objective has been iterated repeatedly until the desired signal has been met.

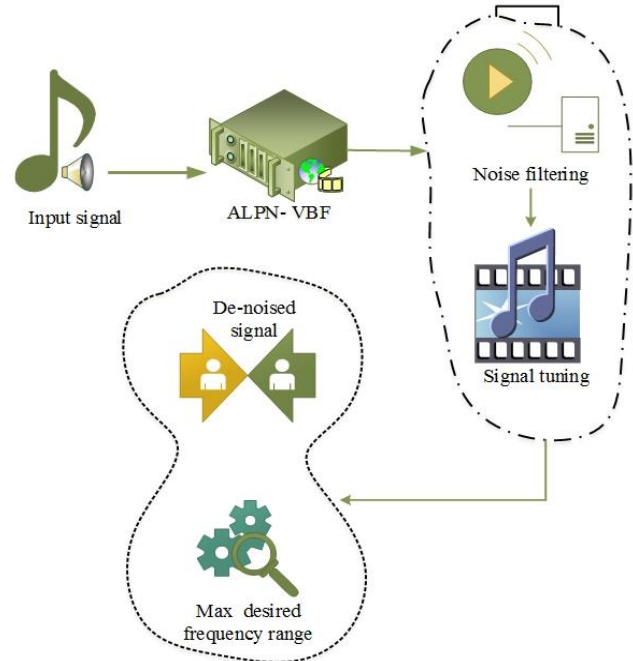


Fig. 2. Proposed architecture

The proposed architecture is detailed in Fig. 2. Finally, the working of the designed efficient VBF has been analyzed by validating the function parameters by applying any voice or sound. Subsequently, the improvement score has been measured by processing the comparison assessment.

### A. Design of ALPN-VBF for HA System

The present work has developed a novel optimized intelligent ALPN-VBF for the HA system to improve the signal clarity of the HA device. Hence, this technique incorporates LPF, HPF, and Ant lion procedures. Usually, the ant lion is inspired by their hunting intelligence by trapping the ant inside the pit. Here, this fitness is utilized to fix the maximum desired frequency and tune the frequency up to the fixed frequency.

$$c = \sum_{n=1}^x p_n 2^{S-n} \quad (1)$$

Here,  $c$  denotes filter coefficients, world length is determined as  $x$ , the required power is represented as  $S$  and signal features are described as  $p$ . Moreover, the  $p_n$  is signal coding that is defined as  $p_n = (-1,0,1)$ . Hence, the filter coefficient is described in (1).

$$F(y) = \|G_d(\omega) - G(\omega, y)\|_2 \quad (2)$$

Here, the response of the zero-phase frequency is denoted as  $G_d$  and  $G(\omega, y)$  is the present noise feature in the input signal. In addition,  $F(y)$  is the objective function signal tuning and filtering. Hence, (2) is utilized to neglect the present noise in the input signal. Moreover, the signal range has been tuned by fixing the maximum signal range and tuning the input signal during the execution of the signal processing (3).

$$A(y) = \max(0, u(y) - e_b) \quad (3)$$

Here,  $e_b$  is the detection parameter to find the equivalent fixed maximum signal range  $u(y)$  and  $A(y)$  is the threshold setting parameter. Now, the minimum frequency of the sound is fixed using (4),

$$\min F(y) = \beta_1 F_1(y) + \beta_2 A(y) \quad (4)$$

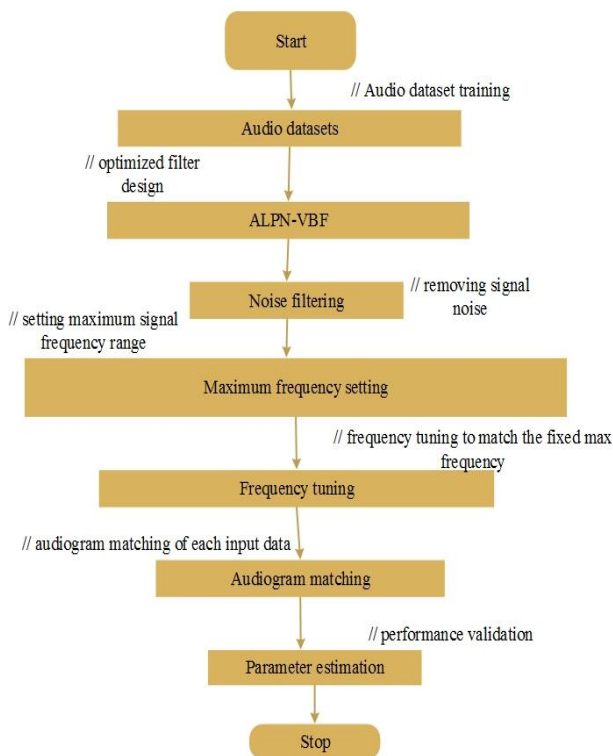


Fig. 3. Flow of novel ALPN-VBF

The design of ALPN-VBF Where,  $\beta_1, \beta_2$  are the signal's weight coefficients. In addition, the system is trained with more different signals; each signal has specific frequency metrics. Hence, the signal selected from the total set of the collected sound signal is processed by (5).

$$\begin{cases} 1 & \text{ifrand} > \max(u(y)) \\ 0 & \text{ifrand} \leq \max(u(y)) \end{cases} \quad (5)$$

Besides, the tuning process of the signal coefficients is detailed by (6), if the statement has been worked with 0 and 1 status. Hence, for the random selection process, 0's and 1's conditions were used.

$$T(s) = \frac{\max(u(y)) + \min(u(y))}{2} \quad (6)$$

$$\text{optimum frequency} = F(y), \text{ if } T(s) < \max(u(y)) \quad (7)$$

During the signal transmission process, the optimal range signal (7) has been obtained then the optimization iteration has been stopped. Otherwise, the iteration has been continued till the desired signal is met.

The function process of the novel ALPN-VBF is executed in the way of Fig. 3. The key function of this designed filter is to filter the noise signal with a highly appropriate range and to tune frequency based on the deaf people hearing capacity.

#### IV. RESULTS AND DISCUSSION

The planned research solution is implemented in the MATLAB framework, running in the windows 10 platform. Initially, benchmark data is collected from the standard sites, it contains several sounds like animals, birds, humans, etc., then a novel ALPN-VBF has been developed with desired tuning parameters. Consequently, the signal executing and frequency tuning performance have functioned.

##### A. Case Study and Performance Validation

To analyse the performance of the designed filter, benchmark data has been utilized. From that bird sound and public transport, sound has been processed to analyze the capacity of the noise filtration function and tuning performance.

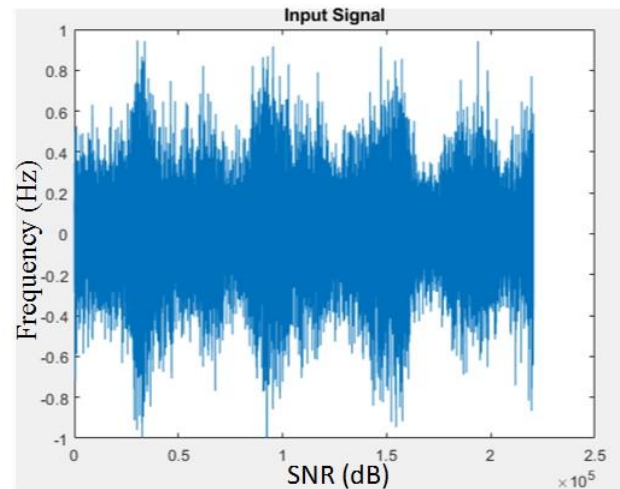


Fig. 4. Input signal frequency

The generated signal plot after importing the sound dataset is described in Fig. 4. The frequency wave of the input signal is varied based on the audio maximum frequency range.

1) *Matching error*: In any filtering concept, finding the signal error or mismatch became the major concern for validating the filter functions. The matching error parameter has been estimated for three audiograms.

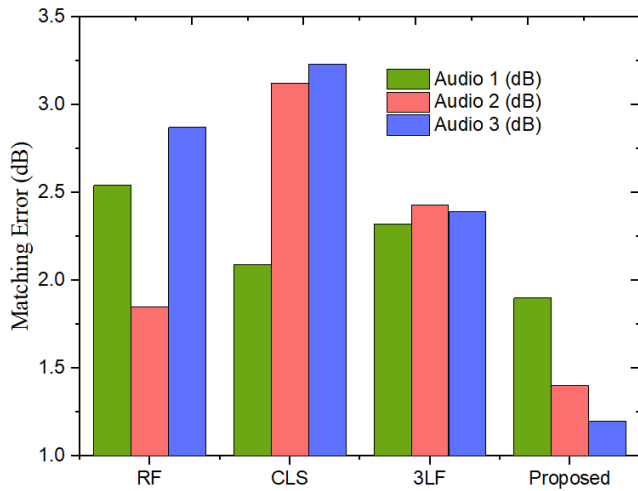


Fig. 5. Matching error assessment

In addition, to find the improvement score of the proposed model, a few existing models have been obtained, such as reconfigurable filter [19], Constrained-least-square (CLS) filter [20], and 3 level Filter (3LF) [21]. Here, the RF has obtained an error score of 2.54 dB for the first audiogram, 1.85dB for the second audiogram, and 2.87 dB for the third audiogram. Hence, evaluating these statistics, the method RF [22] has recorded the average matching error score as 2.4db. Furthermore, the CLS filter has gained the error score for the first audiogram is 2.09dB, 3.12 dB for a second audiogram, and 3.23 dB for the third audiogram. Hence, the recorded average error score by the method CLS [23] is 3dB. In addition, the method 3LF [24] has gained 2.32 dB for the first audiogram, 2.43 dB for the second audiogram, and 2.39 dB for the third audiogram. The average recorded score for matching error is 2.35dB. Considering all those existing models, the proposed approach has obtained 1.9dB for the first audiogram, 1.4dB for the second audiogram, and 1.2dB for the third audiogram. Hence, the obtained average matching error for the proposed model is 1.5dB. The matching error statistics is graphically represented in Fig. 5.

2) *Delay validation:* Measuring the time for the audio frequency transmission is called a delay. Mainly, the filters which have the complex function modules have required more time to execute; this causes time delay.

Moreover, the RF approach has achieved an 18.54 ms delay, the CLS filter has recorded a 7.9ms delay, 3LF has achieved the frequency delay as 12.9ms, and the proposed strategy has recorded the delay score to broadcast the signal frequency is 5ms. The delay statistics are described in Fig. 7.

3) *Power Consumption:* In digital signal application, power is the key metric for transferring the signal, the utilized power by the novel ALPN-VBF has been measured with Watt unit.

In the proposed research, dual filters were used: LPF and HPF, so power is a major concern for analyzing the maximum required power for HA system fitters. Besides, the reason for consuming high power is the poor filtering capacity. So, the method with less power consumption has a high noise removal rate. The CLS filter has gained the power consumption score of

0.05W; RF has obtained the power consumption score of 0.398W, and the proposed approach recorded the consumed power as 0.02W. The measurement of power utilization is described in Fig. 6 and the overall power consumption validation measurement of the proposed ALPN-VBF is tabulated in Table I.

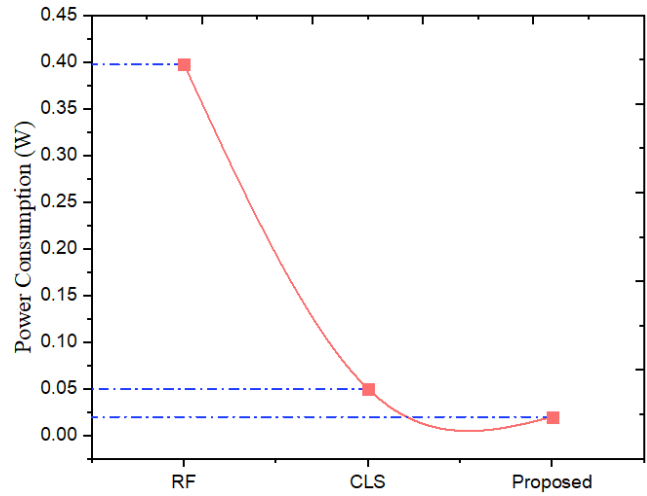


Fig. 6. Power consumption validation

TABLE I. THE OVERALL COMPARISON STATISTICS

comparison statistics					
Methods	Matching error dB			Power consumption (W)	Delay (ms)
	Audio1	Audio 2	Audio 3		
RF	2.54	1.85	2.87	0.398	18.54
CLS	2.09	3.12	3.23	0.05	7.9
3LF	2.32	2.43	2.39	0.03	12.9
Proposed	1.9	1.4	1.2	0.02	5

**B. Discussion**

The designed ALPN-VBF has recorded an outstanding performance from all the comparison assessments than the compared approach. This proved that the involvement of the combined HPF and LPF in one filter bank had provided the finest outcome.

TABLE II. OVERALL PARAMETER VALIDATION OF ALPN-VBF

Overall performance assessment		
Power consumption	0.02 W	
Delay	5ms	
Matching error	Audiogram 1	1.9
	Audiogram 2	1.4
	Audiogram 3	1.2

The overall metrics measurement of the proposed ALPN-VBF is tabulated in Table II. Hence, the designed filter procedure is suitable in HA application to maximize the hearing capacity of deaf people by affording the clarity sound frequency.

## V. CONCLUSION

Improving the performance of the HA application is the most required paradigm for deaf people. Hence, this present work has attempted to design a novel optimized power filter name as ALPN-VBF for the HA system. Initially, the set of audiograms was taken and trained to the system then the filtering process was performed to de-noise the signal. Consequently, the frequency range of the signal is tuned to the maximum desired level with the help of ant lion fitness. Once the desired level frequency has been met, it has transferred to the other end. Finally, the key metrics were calculated and compared with other models. The proposed model has achieved the minimum matching error of 1.5 dB; considering the other models, it has diminished the matching error up to 1%. Besides, the recorded power consumption is 0.02W; compared to other models, it has shown an improvement rate of up to 2%. The delay measure recorded by the designed strategy is 5ms, by comparing other models, has described the improvement rate up to 12%. Hence, the designed model is suitable for the HA system.

## REFERENCES

- [1] P. V. Praveen Sundar, D. Ranjith, T. Karthikeyan, V. Vinoth Kumar, and B. Jeyakumar, "Low power area efficient adaptive FIR filter for hearing aids using distributed arithmetic architecture," *International Journal of Speech Technology*, vol. 23, no. 2, pp. 287-296, 2020.
- [2] S. R. Rammohan, N. Jayashri, M. A. Bivi, C. K. Nayak, and V. R. Niveditha, "High performance hardware design of compressor adder in DA based FIR filters for hearing aids," *International Journal of Speech Technology*, vol. 23, no. 4, pp. 807-814, 2020.
- [3] G. NagaJyothi, and S. Sridevi, "High speed low area OBC DA based decimation filter for hearing aids application," *International Journal of Speech Technology*, vol. 23, no. 1, pp. 111-121, 2020.
- [4] R. Ramya, and S. Moorthi, "Audiogram matching in hearing aid using approximate arithmetic," *Multidimensional Systems and Signal Processing*, vol. 2, no. 4, pp. 1199-1215, 2021.
- [5] A. Elkhoully, A. M. Andrew, H. A. Rahim, N. Abdulaziz, M. Abdulmalek, M. N. M. Yasin, M. Jusoh, T. Sabapathy, and S. Siddique, "A Novel Unsupervised Spectral Clustering for Pure-Tone Audiograms towards Hearing Aid Filter Bank Design and Initial Configurations," *Applied Sciences*, vol. 12, no. 1, pp. 298, 2022.
- [6] N. Shankar, G. S. Bhat, and I. M. S. Panahi, "Real-time dual-channel speech enhancement by VAD assisted MVDR beamformer for hearing aid applications using smartphone," *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2020.
- [7] K. Zaman, and C. Direkoğlu, "Classification of Harmful Noise Signals for Hearing Aid Applications using Spectrogram Images and Convolutional Neural Networks," *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, IEEE, 2020.
- [8] B. J. Borgström, M. S. Brandstein, G. A. Ciccarelli, T. F. Quatieri, and C. J. Smalt, "Speaker separation in realistic noise environments with applications to a cognitively-controlled hearing aid," *Neural Networks*, vol. 140, pp. 136-147, 2021.
- [9] Y. Hioka, K. Kobayashi, K. Niwa, "Improving speech intelligibility using microphones on behind the ear hearing aids," *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSp)*, IEEE, 2019.
- [10] S. Aggarwal, and N. Chugh, "Signal processing techniques for motor imagery brain computer interface: A review," *Array*, vol. 1, pp. 100003, 2019.
- [11] F. Henry, M. Glavin, and E. Jones, "Noise Reduction in Cochlear Implant Signal Processing: A Review and Recent Developments," *IEEE reviews in biomedical engineering*, 2021.
- [12] W. Wang, G. Zhang, L. Yang, V. S. Balaji, V. Elamaran, and N. Arunkumar, "Revisiting signal processing with spectrogram analysis on EEG, ECG and speech signals," *Future Generation Computer Systems*, vol. 98, pp. 227-232, 2019.
- [13] G. Allwood, X. Du, K. M. Webberley, A. Osseiran, and B. J. Marshall, "Advances in acoustic signal processing techniques for enhanced bowel sound analysis," *IEEE reviews in biomedical engineering*, vol. 12, pp. 240-253, 2018.
- [14] Y. Ren, Z. Jiang, and V. Van, "A General Variable Bandwidth Microring Filter for Lossless Bandwidth Tuning," *Journal of Lightwave Technology*, 2021.
- [15] K. A. Swamy, and Z. C. Alex, "Efficient low delay reconfigurable filter bank using parallel structure for hearing aid applications with IoT," *Personal and Ubiquitous Computing*, pp. 1-14, 2021.
- [16] V. R. Niveditha, S. Palaniappan, K. Naresh, C. K. Nayak, and B. Swapna, "High speed low area decimation filter for hearing aid application," *International Journal of Speech Technology*, pp. 1-7, 2021.
- [17] T. Ma, Y. Wei, and X. Lou, "Modularized Design of Completely Reconfigurable Nonuniform Filter Bank for Hearing Aid Systems," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021.
- [18] S. C. Marcum, E. M. Picou, T. Steffens, R. Hannemann, V. Vielsmeier, M. Schecklmann, B. Langguth, and W. Schlee, "Conventional versus notch filter amplification for the treatment of tinnitus in adults with mild-to-moderate hearing loss," *Progress in brain research*, Elsevier, 2021, pp. 235-252.
- [19] T. Devis, and M. Manuel, "Hardware-efficient auto-reconfigurable hearing aids using 3-level octave interpolated filters for auditory compensation applications," *Physical and Engineering Sciences in Medicine*, pp. 1-14, 2021.
- [20] V. V. Mahesh, and T. K. Shahana, "Design and synthesis of filter bank structures based on low order constrained least square and minimum phase methods for audiogram matching in digital hearing aids," *Health and Technology*, vol. 11, no. 1, pp. 153-168, 2021.
- [21] T. Devis, and M. Manuel, "A low-complexity 3-level filter bank design for effective restoration of audibility in digital hearing aids," *Biomedical Engineering Letters*, vol. 10, no. 4, pp. 593-601, 2020.
- [22] U. S. Rawandale, S. R. Ganorkar, and M. T. Kolte, "Audiogram Study in Filter Bank Used for Hearing Aid System to Enhance the Performance," *20226th International Conference On Computing, Communication, Control And Automation (ICCUBEA, Pune, India, 2022)*, pp. 1-4, doi: 10.1109/ICCUBEA54992.2022.10010832.
- [23] U. S. Rawandale, and M. T. Kolte, "Design, Development and Analysis of Variable Bandwidth Filter Bank for Enhancing the Performance of Hearing Aid System," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 6, pp. 391-401, 2021, DOI: 10.22266/ijies2021.1231.35
- [24] U. S. Rawandale, and M. T. Kolte, "Study of Audiogram for Speech Processing in Hearing Aid System," *2019 IEEE Pune Section International Conference (PuneCon)*, IEEE 2019, DOI: 10.1109/PuneCon46936.2019.9105706.

# Performance Evaluation of Photovoltaic Projects in Latin America

Cristian León-Ospina, Heyner Arias-Zarate, Cesar Hernandez  
Universidad Distrital Francisco José de Caldas, Bogotá D.C, Colombia

**Abstract**—Photovoltaic solar energy has been booming worldwide due to the scarcity of non-renewable resources, from there arises the need to modernize and innovate in the methodologies for the use of energy resources as well as in the correct installation of this system at an urban or level rural. In recent decades in Latin America there have been presented advances in the implementation of photovoltaic projects, which is why this document aims to evaluate the feasibility of these at a technical and technological level, so that they are in accordance with the systems implemented in Asia, Europe, and North America. The analysis determined that the main factors affecting the feasibility of a photovoltaic execution project are economic and technological, in addition to the adverse impacts found on the ecosystem and the local population, but in general it was observed that they are weaknesses that can be corrected, since there are different countries that are working on establishing strategies to educate their community, so there is an improvement in the quality of life in sectors with high CO<sub>2</sub> pollution and lack of fossil fuels.

**Keywords**—Evaluation; Latin America; photovoltaic; project

## I. INTRODUCTION

At present, due to the inconveniences presented by the scarcity of non-renewable resources that are used for energy generation in the world, it is necessary to implement new methods that take advantage of natural resources, as is the case of photovoltaic systems. In this procedure, solar energy is harnessed by storing, regulating, and transforming it into electricity for use in residential and industrial electrical installations [1].

The definition of solar energy refers to the electromagnetic radiation emitted by the sun that can be harnessed using multiple technologies. Modeling and simulation of these types of systems have become a widespread practice in the academic and industrial world [2]. They are essential tools for the design and installation of a solar plant, and for the estimation of its productivity.

Photovoltaic systems are recognized as a group of components that make it possible to harness and use electromagnetic radiation emitted by the sun to produce electricity. These types of systems are identified according to the types that compose them and the materials or elements with which they are manufactured and put into operation.

This document details the analysis of the factors that make it possible for a photovoltaic project to be carried out satisfactorily in a sustainable, energetic, and economic way. This analysis aims to contemplate the feasibility of these projects and contribute to the strengthening of knowledge

about photovoltaic energy and therefore motivate research and development of strategies that optimize it, to contribute to the improvement in the quality of life and in turn with the reduction of CO<sub>2</sub>.

This was done through specific IEEE bibliographic sources that document the axis of analysis of this project; these documents were compared by analyzing the different implementation strategies of photovoltaic projects in specific Latin American countries, using graphs, figures, and descriptions of each. To contemplate the factors those were most relevant when materializing this type of project.

This document brings a unique added value as it offers innovation and solutions for energy problems that are difficult to address. This means that this document stands out from other documents because of the usefulness of its content. Photovoltaic technology is especially useful for companies and organizations that seek to optimize production and service processes for communities with difficulties in accessing electricity. This makes it possible to save time and money by providing information that facilitates decision-making on critical factors of the project and at the same time improves the efficiency of the photovoltaic generation process.

## II. RELATED WORKS

In [3] performs a return analysis on the study of various photovoltaic systems, investigating three generations of photovoltaic technology taking into account efficiency, embodied energy and return on investment in energy to identify relevant elements that allow generating technological development. One of the main objectives is to find ways to reduce the financial costs of the production of photovoltaic systems. It is found that the first-generation wafer-based technology has a higher efficiency than the second-generation thin film, but at a higher cost per unit area and that for the third generation the technologies have not sufficiently emerged yet. But it is also highlighted that energy performance and environmental impacts or greenhouse gas emissions must be taken into account, which can be indicators of the benefits and costs, this was found that the thin film has the best advances in performance energetic and is currently working better [3].

In [4] described the impact assessment of net metering for Distributed Residential Photovoltaic Generation in Peru. The economic impact of what is known as net metering of the photovoltaic system interconnected to the grid is analyzed, where the user is able to deliver or receive energy from the grid depending on whether there is an energy surplus or not

receive missing energy from the network to meet their needs through bidirectional meters, in addition, an analysis of the legal and regulatory mechanisms that define as such the economic conditions of energy exchange with the electrical distribution network in the city of Arequipa is carried out. This basically provokes a discussion in terms of the advantages and disadvantages that both distribution companies and end users present, since the payback time of applying this methodology is quite long and the effectiveness depends on the geographical area. Because if the location has a sufficient level of solar energy, it is more likely that the investment will be recovered a little faster. On the other hand, the distribution companies would see their business affected since it depends on the end user's consumption being higher and not lower due to the injection of energy that they would generate on the grid, for which the prices in the market would be affected [4].

In [5] presents an analysis of the different subsidies provided by the Mexican government to those who promote projects that implement renewable energy, more specifically photovoltaic energy, taking advantage of the great growth that this has had in recent years, measuring the economic feasibility of said generation projects. will have on the energy market in the country, sustaining that the use of this type of technology mitigates climate change, that this type of energy does not generate greenhouse gases and is booming, so Mexico is taking advantage of this opportunity to strengthen the agricultural sector, taking into account that Mexican investors carefully evaluate the risks of contributing their capital for this type of project. In order to see what factors must be taken into account, a study is also made of the dates of different regions of the country to see how much solar energy could be counted on to consider the energy potential, and what investment mechanisms can be used taking into account the affectation that this will have to implement in the agricultural sector due to the strong initial investment that must be made, so the incidence of subsidies on the viability of these projects will also be evaluated due to the analysis of the cost of the energy generated and the return time of the investment, but according to the analyzes carried out, everything points to the fact that without these subsidies, due to the high costs, this type of generation would not be profitable [5].

### III. PHOTOVOLTAIC PROJECTS IN LATIN AMERICA

#### A. Scientific Databases

When delving into the research of photovoltaic projects in Latin America, it was used as query tools several databases among which IEEE Xplore and Scopus stand out. For this case IEEE publications that have been made on this topic in the last 20 years were used.

1) *Documents per year*: In this section we will recognize the evolution of research on issues related to photovoltaic systems in Latin America in the last 20 years according to the information provided by the databases. See Fig. 1.

2) *Documents by region*: The countries that have done the most research in the implementation of photovoltaic systems in this part of the continent are Brazil (87), Uruguay (61), Peru (38) and Ecuador (15).

3) *Types of documents*: It is highlighted that the documents issued by educational or academic institutions are conferences (228), scientific articles (123) and journals (3), generating a total of 354 documents.

4) *Most frequent keywords*: Based on the bibliometric analysis it was carried out using several tools such as VOS viewer, after was mapped for this case the keywords that are most used in the documents found the IEEE databases, thus recognizing which are the focuses of the research carried out in the area.

Fig. 2 shows the topics most frequently mentioned in the documents are solar energy, photovoltaic cells, photovoltaic systems, all implemented in Latin America.

#### B. Projects with the Highest Energy Production in Latin America

In this area of the continent the use of renewable energies is booming, in this case photovoltaic solar energy in recent years has excelled dramatically in several countries, some of the largest producers are: Brazil, Mexico, Chile and Argentina.



Fig. 1. Research trend in recent years.

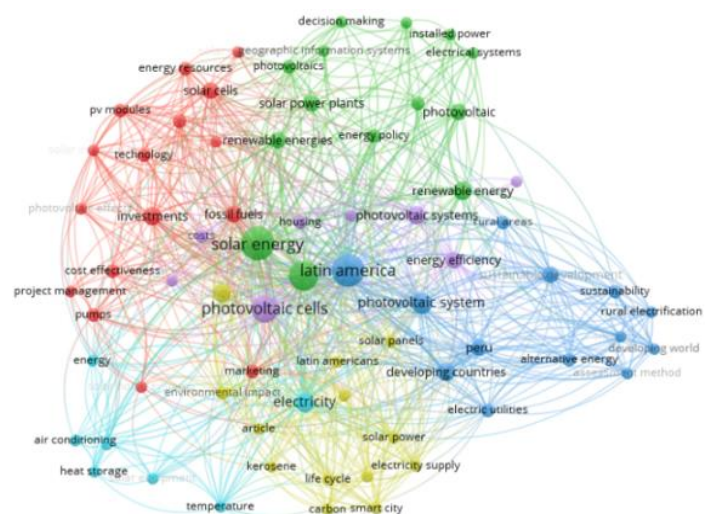


Fig. 2. Visualization of the network related to the keywords covered in the articles referring to photovoltaic systems.

Each of these nations was delved into to identify their most emblematic projects in this type of energy technologies, thus recognizing their technical specifications in order to demonstrate the current energy production and its contribution to the national interconnected system.

1) *Brazil*: It is the country with the largest population area in Latin America, because of this the increase in energy demand extends annually that is why they are implementing the use of new technologies for energy generation, among those are the photovoltaic systems who in recent years it has increased its power generation capacity exponentially as shows the Fig. 3.

Recognizing this renewable energy generation that the country has had, the attention will focus on the exploration of some of the largest solar farm projects that the country currently has, among which are:

a) *Ituverava solar park*: This solar park has been operating since 2017, being the largest plants in South America, has an approximate area of 600 hectares, houses around 850,000 solar panels, with a maximum production capacity of 550 GWh per year [7]. It is located in the state of Bahia in the municipality of Tabocas do Brejo Velho.

In the area where this solar park was built, the presence of a lot of biodiversity is evident, which is why mechanisms have been implemented to mitigate the environmental impact such as artificial nests for birds, recycling and water treatment systems and a mini solar farm. This is intended to supply a part of electricity to a neighboring community. On the other hand, the investment made by Enel for the construction and execution of this plant was around 4 billion dollars, which is intended to supply the energy consumption of around 268,000 homes, thus preventing the emission of approximately 318,000 tons of CO<sub>2</sub> [7].

b) *Nova olinda solar park*: It is currently one of the largest solar photovoltaic power generation parks in Latin America, its approximate installation area is 690 hectares, it has a total installed capacity of 292 MW, producing approximately a maximum capacity of 600 GWh [8]. It is located in the state of Piauí in the municipality of Ribeira do Piauí.

The company managing the service invested around 300 million dollars in the construction of nova Olinda, thus maximizing the energy coverage of around 300,000 homes, generating a reduction in greenhouse gas emissions of around 350,000 tons.

Both solar parks are managed by Enel, an Italian company that is one of the world's largest providers of renewable energy optimization for the purpose of cleanly electrifying the countries in which it operates.

2) *Mexico*: This Central American country has the second-best population in photovoltaic generation, due to the global energy crisis that has been presenting, Mexico has seen the need to implement new generation parks in order to meet the demand it currently has. The evolution it has had in the photovoltaic generation capacity in the last decade has grown

exponentially, amplifying its generation in the last 5 years. Fig. 4 shows the trend of this energy production.

Next, some of the largest photovoltaic solar generation parks in the country will be observed.

a) *Villanueva solar power plant*: Since 2018 the Villanueva solar photovoltaic plant has been in operation with an installation area of approximately 2,400 hectares, thus projecting a production of 1700 GWh per year when it is fully operational. There are around 200 to 300 of solar panels [9], this plant is located in the state of Coahuila in the town of Viesca.

The investment for this plant was US\$650 million, with a strong focus on environmental impact, waste management, CO<sub>2</sub> emission reduction and economic growth for the communities living near the solar plant.

b) *Magdalena II solar power plant*: In 2019, the construction of this work was completed, housing an area of approximately 800 hectares, at maximum capacity it could generate 600 GWh per year. [10].

One of the technological advances that this plant has is that its photovoltaic modules are bifacial, in order to take advantage of the incidence of the sun on both sides of the solar panel [10], and these panels have a monitoring mechanism which is programmed to obtain the greatest use of solar radiation.

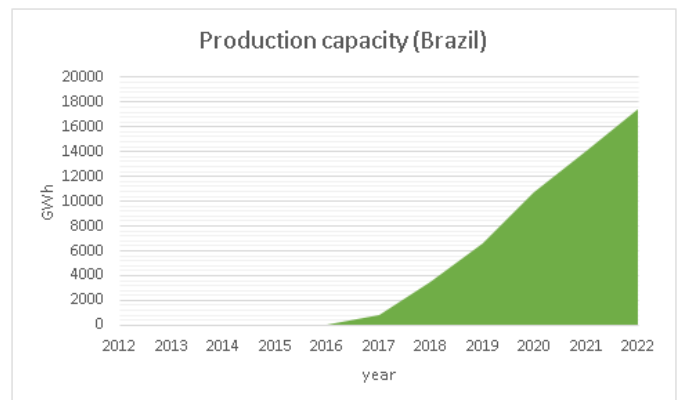


Fig. 3. Photovoltaic solar production capacity in Brazil [6].

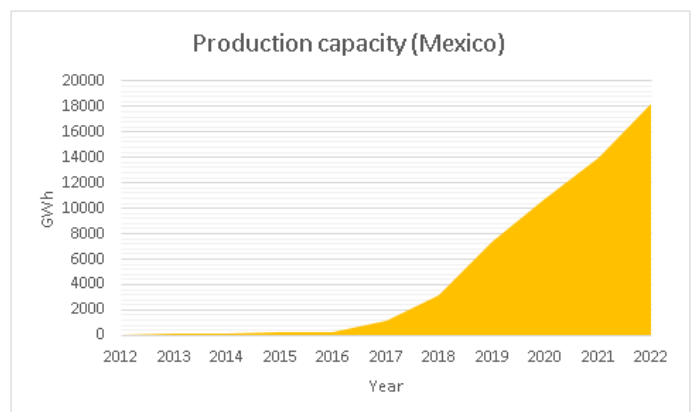


Fig. 4. Photovoltaic solar production capacity in Mexico [6].



This Park is located in the state of Tlaxcala and is intended to supply the consumption needs of approximately 456,000 Mexican households.

3) *Chile*: In this area of the continent there is the best solar radiation in the world, which is why the boom it has acquired in the last decade on the generation of photovoltaic solar energy has been wide.

It must recognize that lately the materials that are being used by the entities that execute this type of energy projects are of high efficiency, this in order to make the most of the energy resource. Fig. 5 shows the increase in the production capacity of gigawatt hours of photovoltaic solar energy.

Some of the projects that stand out most in the country are the following:

a) *Domeyko solar plant*: The construction of this project is in charge of Enel Power Chile, it has an approximate area of 700 hectares where approximately 473000 bifacial photovoltaic modules are installed [6], in order to generate greater use of the reception of solar radiation, It should be noted that in this area of the world is where there is greater energy efficiency regarding this type of energy. When the plant operates at its maximum capacity it can generate approximately 590 GWh per year. It is located in the region of Antofagasta.

The cost of this project was 164 million dollars, the execution of this was carried out during the pandemic therefore its workers complied with all safety protocols in order to guarantee their health and that of their community, this project aims to avoid the emission of CO<sub>2</sub> of approximately 439000 tons annually [6].

b) *Sol de lila photovoltaic park*: This solar plant is located in the Atacama Desert, has an installation area of approximately 720 hectares, where approximately 408,000 bifacial photovoltaic modules are installed [12]. when the maximum operation is found, it is expected to generate approximately 500 GWh per year, thus benefiting the surrounding population of the Antofagasta region.

The greenhouse gases that can be mitigated when this solar park is in execution would be approximately 372000 tones. The execution of this work was carried out by more or less 500 workers in times of pandemic, each of them maintaining safety protocols.

4) *Argentina*: In this country, despite being in an area where there is a great solar radiation, it had not invested much money in energy innovation due to its political and economic problems.

Despite the situation, it can be seen that in the last decade it has begun to increase its photovoltaic solar energy production capacity gradually since 2017, thus recognizing its need to implement renewable energies in the national interconnected system. Fig. 6 will show the increase in energy production.

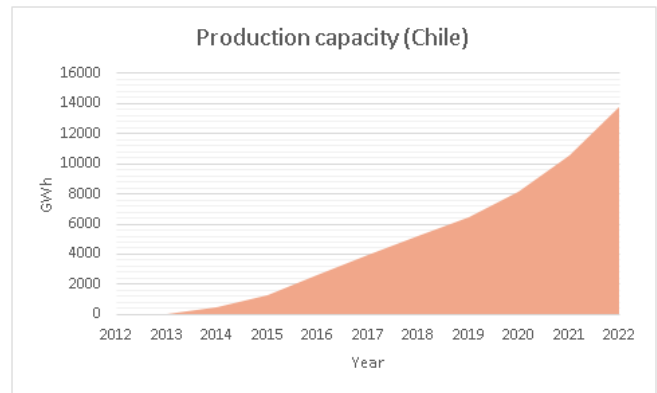


Fig. 5. Photovoltaic solar production capacity in Chile [11].

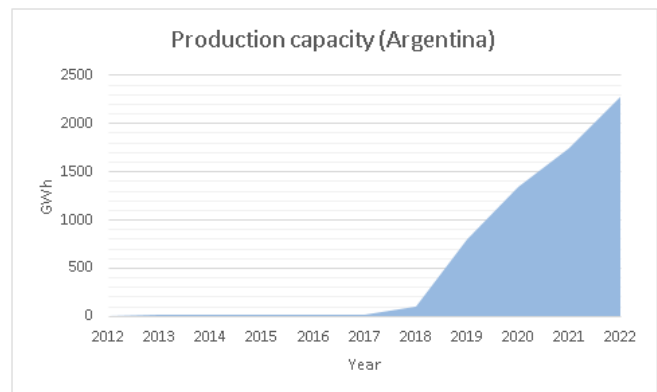


Fig. 6. Photovoltaic solar production capacity in Argentina [11].

The most relevant solar plants in the country are:

a) *Caucharí solar plant*: The energy company Jemse started up the largest solar energy plant in Argentina, located in the Puna region, with an area of approximately 800 hectares and an installation of 90,000 solar panels, producing 215 GWh annually at maximum capacity [13].

The power that handles this solar plant is 300 megawatts, the implementation cost was 390 million dollars, pretending to cover the energy need of approximately 160,000 homes, it is located at 4020 m above sea level.

b) *Altiplano 200 photovoltaic solar park*: This solar park is located in the town of Olacapato, has an installation area of approximately 350 hectares, locating there a total of 500550 solar panels, producing at maximum capacity a power of 200 MW [14]. The investment for this project was 313 million dollars, this project was created by merging the projects of Puna and Altiplano. The project is managed by Cammesa for a period of 20 years.

#### IV. TECHNICAL AND ECONOMIC FEASIBILITY OF PHOTOVOLTAIC PROJECTS

##### A. Type of Environment or Geographic Area

The optimal environment to implement a photovoltaic system is one that allows the greatest solar radiation throughout the day, although this depends on some atmospheric phenomena that are listed in Table I.

TABLE I. ATMOSPHERIC PHENOMENA PRESENTED IN SOLAR RADIATION [25]

Phenomenon	Definition
<b>Absorption</b>	A large accumulation of solar radiation that is filtered by the gases present in the atmosphere thus preventing its transit to the surface of the earth.
<b>Diffraction</b>	It is a phenomenon that occurs when solar radiation transits through clouds, the incident ray when reaching these is fragmented in such a way that each of these has less energy.
<b>Dispersion</b>	For this case, the sun's rays are deflected by dust particles water molecules present, thus preventing their arrival on earth.
<b>Reflection</b>	This phenomenon has the characteristic that when it meets the particles that are suspended in the clouds, as well as the gases that it contains, they cause the sun's rays to be returned to space.

Complemented the above, it is necessary to emphasize that this energy resource has great variability also due to climatic factors such as the conditions that occur throughout the day, as well as the change at night and the time of year in which this system is used.

Photovoltaic installation depends on the place where it can be located, if they are generation parks at a macro level such as those managed by large energy companies in the world it is necessary that this is located in a desert or arid area allowing ease of construction and in turn the feasibility of the system mitigating its environmental impact, on the other hand if our photovoltaic installation is micro level this can be located anywhere in the world that is outdoors and can capture solar radiation.

Depending on the location it has at its latitude this will define an estimate for the angle of inclination that solar panels can possess if they are of fixed type. Fig. 7 shows a mapping of the areas where there is greater solar radiation in the world.

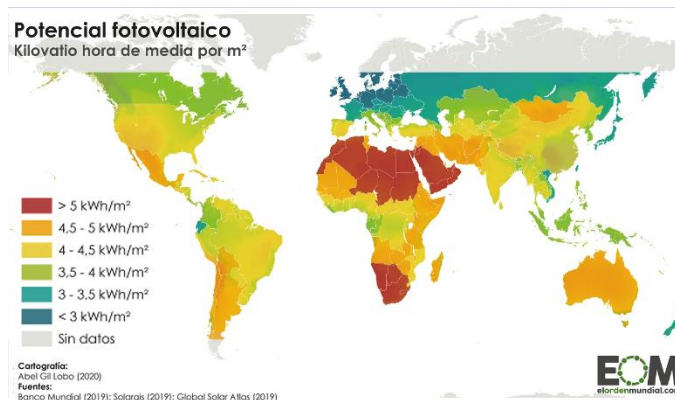


Fig. 7. World map of the areas with the greatest solar irradiation.

### B. Space required for the Installation of these Systems

1) *Solar parks*: When large photovoltaic projects are executed, it is necessary that these are located in areas where the greatest irradiation in the country is present, such as deserts, among others, this type of spaces must have a level ground so that when installing the structure of the solar panels there are no variations in the This would affect the optimal reception of this energy resource.

2) *Isolated photovoltaic systems*: This type of installation is intended to locate the solar panels in a clear area either located on the ground or attached to the roof of the place where they are going to be implemented, this in order to avoid deviations or obstructions of the solar radiation that is directed to the panels. For the rest of the components such as the inverter, batteries, regulator and protection systems, it is required that these be located in a place that is not outdoors and in turn allows a high flow of air in order to reduce the temperature of these electronic devices.

3) *Grid-connected PV systems*: When they are installed at the residential level you can follow the parameters seen in the isolated systems, there are other cases that are being presented at an industrial and commercial level in the continent are the integration of solar modules on the facades and terraces of buildings and factories in order to take advantage of the great irradiation that it reaches this type of structures and likewise contrasting with the architecture of these so that they do not lose their attractiveness. Some grid operators are encouraging this type of microgrid and are regulating its correct installation to minimize dependence on other types of non-renewable energy sources [17].

Some countries such as Brazil are conducting multiple investigations that are allowing to evaluate the energy performance of this type of systems, which is why certain scenarios propose to use several types of inclinations of solar modules in order to recognize what is the optimal angle for the reception of solar radiation , P Grams such as PVsyst allow us to model some of these scenarios to have great certainty about the nominal power that can be obtained in a period of time, thus corroborating what is calculated in the theory.

### C. Required Materials

In this section, attention is focused on the materials used for installation in a given site or area.

1) *Panel brackets*: The fastening systems aim to support the solar modules as well as to fix the inclination they must carry in order to receive as much of the energy resource or as possible. There are two types of supports which are fixed supports and variable supports. See Fig. 8.

2) *Solar modules*: This material is one of the most important for this type of energy generation systems, initially the panels were single-facial type thus allowing the reception of energy radiation only on one side, currently a new module has been implemented in the market that takes advantage of both sides called bifacial module as mentioned above allows the capture of solar radiation in the upper part and lower of it. In distributed generation it can be seen that there is a generation power limit of 5 MW as it happens in some Latin American countries, due to this restriction there has been the need to optimize energy capture in order to make the most of the implementation of these systems [18].

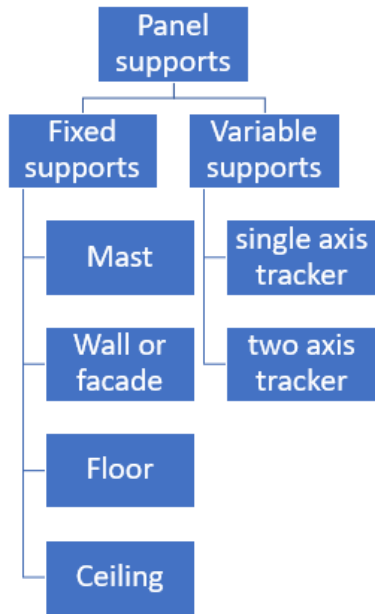


Fig. 8. Types of supports for solar panels [25].

3) *Cables*: For the correct installation of this type of systems it is necessary to implement electrical conductors that are dimensioned in such a way that they comply with the maximum admissible parameters according to the type and installation and that in turn is governed by the rules or regulations that it administers in each country as appropriate. Some of the characteristics of these elements will be summarized below in Table II.

TABLE II. TECHNICAL CHARACTERISTICS OF DRIVERS

Feature	Details
<b>Weather resistance</b>	Maximum conductor temperature resistance to external temperatures, resistance to UV rays, etc.
<b>Mechanical strength</b>	Impact, abrasion and tear resistant
<b>Mitigation environmental impact</b>	Low emission of corrosive gases, no fire spreader, halogen-free, etc.
<b>Life</b>	Endure a minimum of 30 years.

4) *Connectors*: The main characteristic of this type of elements is that they allow the connection of electrical conductors without the need for a splice and in turn have resistance to weathering, thus facilitating the protection of the integrity of the people who transit through the area or manipulate these connections at times of the year when there is high humidity or weather conditions are not the most favorable for the photovoltaic system.

5) *Electrical connection system*: To generate the control and correct distribution of the photovoltaic system and also of its protection elements such as thermomagnetic switches, commissioning systems and the distribution of the branch circuits contemplated for residential or industrial use, it is necessary to do it in a connection box. Among its characteristics is that it must be covered and preferably not outdoors, this in order to prevent the risks of an electric shock

when it occurs in a humid environment, the thermomagnetic switches or fuses that this connection box carries must comply with the electrical regulations of each country thus guaranteeing a maximum power support without risk to a short circuit or Overvoltage, CAD to element that is connected there must be labeled so that it can be clearly evidenced to which branch circuit it corresponds and the electrical conductors must comply with the color code presented by the local regulations.

D. Investment

Depending on the use that will be given to the photovoltaic installation, the influence of some evaluation criteria must be taken into account that allows the profitability of the projects to be recognized.

1) *Evaluation criteria*: Levelized cost of energy (LCOE): is a parameter that allows us to express the cost of a power generation system in a time interval. The variables that fall under this parameter are the net costs of the investment, the rate established for the discount and the total energy used. A project is viable if the LCOE has a lower value than the energy cost incurred in that investment. The expression defined for this case as shown in equation 1.

$$LCOE = \frac{I_0 + \sum_{t=1}^n \frac{A_t}{(1+i)^t}}{\sum_{t=1}^n \frac{M_{el}}{(1+i)^t}} \quad (1)$$

Internal rate of return (IRR): is a discount rate that generates a balance between expenses and income, thus causing the net present value to be zero. There is profitability in a project if the internal rate of return is greater than the profitability established for the investment, the expression is defined as shown in equation 2.

$$TIR = \sum_{T=0}^n \frac{F_n}{(1+i)^n} = 0 \quad (2)$$

Net present value (NPV or NPV): it is a parameter that makes the addition of cash flows in relation to the rate of capital cost, thus considers that a project has profitability at an economic level if the value obtained is greater equal to zero, the mathematical expression is shown in the equation 3.

$$VAN = \sum_{t=1}^n \frac{F_t}{(1+k)^t} - I_0 \quad (3)$$

2) *Smaller-scale investments*: Unfortunately investing in the implementation of this type of systems locally such as installing panels in a house can be relatively expensive, these values vary according to the place where the installation materials are purchased, the qualified, certified labor, regulation, and commissioning of the system. It foresees that this type of investment has a payback period of 5 to 15 years, although some countries are allowing that, if this type of systems are connected to the grid, they can then reduce the cost of the kWh that the network operator can charges thus reducing the recovery time.

3) *Larger-scale investments*: Large companies worldwide such as Enel Green Power, are pioneers in the implementation

of this type of systems and in the last decade have generated a large investment in unexplored areas in some Latin American countries, on average investments per project are around 250 million dollars [19]. Based on the type of investment, the company agrees with local governments to manage this energy system for a minimum of 15 years, thus allowing control of it, gradually impacting the percentage of energy generation in each country and guaranteeing the recovery of its investment in a short time.

#### E. Incentives

1) *Paid*: For small solar farms, there are economic rewards for using clean technologies in their electrical installations, these are reflected in the reduction of the cost of kWh and in some cases in the financing of the installation of this type of systems [16].

Companies that manage the largest solar projects in the region receive incentives for mitigating environmental impacts in the area where the project is executed, as well as for the amount of energy they can supply during peak hours, where there is high energy demand.

2) *Unpaid*: Unpaid incentives are those that are presented as support for humanity, such as the brief mitigation of greenhouse gas emissions, reduction in the generation of hazardous substances to the environment and support for the vegetation of the areas where this type of energy systems is implemented.

#### F. Environmental Impacts

1) *Soil*: Despite considering this technology as clean, the environmental impact cannot be completely mitigated since in the case of an installation of a solar park an average of 2 hectares per megawatt installed is required, this over time generates an alteration in the soil thus impacting the habitat that is around this installation, It should be noted that this only applies in areas where there is high vegetation and fauna, therefore a solution mechanism is to execute this type of projects in desert areas.

2) *Atmosphere*: A considerable amount of the environmental impacts that occur when generating photovoltaic solar energy are the presence of aerosols in the atmosphere, these prevent the efficiency of solar radiation therefore it is integral to carry out studies that allow us to

recognize this type of conjunctures and who that is can generate some type of solution in the short and medium term. [15].

3) *Impact on local flora and fauna*: Another type of environmental impacts that occur in this type of systems is when the flora and fauna of the area where it is going to be implemented is affected, as a result of this some projects have not been viable due to the damage they can generate to the population, in other places it has been possible to generate a common agreement between the population and the entrepreneurs with the in order to receive some series of incentives for allowing the construction of this type of works. [26].

## V. COMPARATIVE ANALYSIS OF PROJECTS

Recognizing the great importance that the implementation of large solar parks generates in a country, it is then possible to deepen this exploration in order to demonstrate what characteristics and parameters stand out between each of the projects, such as strengths, weaknesses, environmental impact, energy generated, investment costs and geographical location chosen for this document.

### A. Comparison between Projects in the Same Country

1) *Brazil*: Between Ponta Pora and Fortaleza is the largest area that produces the largest amount of photovoltaic potential in the country, currently Brazil is one of the largest generators of electricity by photovoltaic solar parks. In 2021, it was consolidated among the ranking of the best country is worldwide occupying box 13 [20], with an installed capacity of approximately 14 GW that are currently in operation, in the map of photovoltaic power that Brazil has, it can see in which areas are more prone to the installation of solar parks to ensure that the reception of the resource is optimized.

Recognizing this energy impact that is presenting worldwide, it is necessary to delve into the main characteristics of the most representative solar parks of this country, this in order to have a base that allows us to recognize what parameters of owning a large-scale photovoltaic solar installation in this part of the continent.

As mentioned in Section II, the solar parks of Nova Olinda and Ituverava are those that present at the moment the greatest capacity of energy generation, therefore, each of its primary particularities is synthesized in Table III.

TABLE III. COMPARISON BETWEEN THE PROJECTS CHOSEN

Variables	Brazil		Mexico		Chile		Argentina	
	New Olinda	Ituverava	Villanueva Solar Power Plant	Central solar Magdalena II	Domeyko Solar Plant	Sol de Lila photovoltaic park	Caucharí Solar Park	Altiplano 200 Solar Park
<b>Strengths</b>	They established a dialogue in common agreement with the surrounding indigenous communities and managed to generate an agreement in which they would provide work and educational training to a large part of the inhabitants of this area, this in order to generate social inclusion so that both parties are benefited from the execution of this project.	High reduction of carbon dioxide per year. Development of artificial nests to protect local fauna. Implementation of circular economy for the rational use of water.	Implementation of bifacial panels. CO2 reduction	Implementation of bifacial panels. CO2 reduction	Totake advantage of the energy resource since this plant was located in a desert. Use of bifacial panels.	Totake advantage of the energy resource since this plant was located in a desert. Use of bifacial panels.	State-of-the-art technology in solar panels and inverters.	Union of several distributed generation companies.
<b>Weaknesses</b>	Impact on the common evolution of the indigenous population the Quilombos.	Due to its location, it is predicted that in later years it will affect the soil and its rich production.	Effects on the local population in the cultivation of melon and grass.	Limitations in planting local growers	Limitation in its capacity.	Limitation in its capacity.	Effects on the local population.	Soil damage
<b>Environmental impact</b>	Medium-term damage to the soil in the area.	Effects on bamboo crops and endemic species.	EGO has implemented the sustainable construction site model in order to mitigate environmental impacts.	To mitigate the impact, water saving and waste recycling systems are implemented	Reduction of 439.000 tons of CO2	Reduction of 365.000 tons of CO2	Smog reduction annually	Limitation of CO2 emissions
<b>Potential [MV]</b>	292	254	754	220	204	161	215	200
<b>Location</b>	Piauí	Bay	Coahuila	Tlaxcala	Antofagasta	Atacama	Puna	Olacapato
<b>Cost [USD]</b>	\$ 300.000.000	\$ 400.000.000	\$ 650.000.000	\$ 165.000.000	\$ 164.000.000	\$ 135.000.000	\$ 555.000.000	\$ 313.000.000
<b>System Losses [MW]</b>	73	63.5	188.5	55	51	40.25	53.75	50

Based on this comparison, the magnitude of the investment in this energy technology is recognized, and how it can provide a better quality of life around 600,000 inhabitants of the surrounding areas, although it must be recognized if it generates certain environmental and social impacts that are not completely supplied, such as an affectation to the unique species of that region that despite the efforts to mitigate this impact will never be close to 0%, hence Ituverava presents a higher economic investment despite the fact that its generation potential is lower than that of nova Olinda. It should be noted that in most of its actions it is possible to recognize the great change that they are allowing the country to transcend at the technological and energy level in the medium term in order to mitigate the scarcity of non-renewable resources that is coming by leaps and bounds.

2) *Mexico*: This country has the characteristic of having a limitation in humidity and its vegetation is not very wide, which is why solar radiation is very large especially in the northern part of the country. Recognizing this favorability of the area and based on the limitations of fossil resources that this country can generate; it has been involved in the implementation of renewable energies in order to meet the energy demand [27]. Many of its technologies are similar to those it manages in the United States, all this based on a new law called the energy transition law that was published in 2015 which regulates the growth of clean energy in the national interconnected system.

Based on the above, entities such as Asolmex are at the forefront of irregularly enacting all projects related to solar energy, companies such as Enel have come to this part of the continent in order to generate investments that allow the

execution of high-quality projects with a high capacity of energy generation [21], Next, the characteristics of some of them are shown in Table III.

Many of the characteristics of these large photovoltaic projects are recognized. This multinational has great campaigns to mitigate the environmental impact and in turn generates studies that have also allowed it to reduce the social impact with the surrounding communities implementing in these centers of academic and labor training, as well as the generations of direct jobs with the company for the short and medium term [22].

3) *Chile*: Currently in the country is inserted in the national interconnected system about 25% of energy from photovoltaic solar parks, entering around 6500 MW to the system. In the country Several companies that own solar power plants, some of the most prominent are the following:

- Enel: 18 solar power plants.
- Engie: 9 solar power plants
- Prime energy: 14 solar power plants
- Colbún: 6 solar power plants

Some of the regions with the highest solar radiation and that have the constructions of this type of plants are Antofagasta, Atacama and Metropolitan [23].

Some of the characteristics of this type of energy in the country is that the energy is cheaper can mitigate the environmental impact is available anywhere in the country, its maintenance is easy, and energy can be supplied in places where the power line does not reach. To further complement the importance of the implementation of these solar power plants will proceed to recognize some of the characteristics in Table III that handle the solar plants mentioned in Section II.

It could appreciate that both solar parks have similar characteristics since they are located in a desert, thus taking advantage of the incidence of the sun in this area limiting its environmental impact with the flower and fauna of this country. Among its policies for mitigating the environmental impact are the use of circular economy for the rational use of water, recycling solid waste and the limitation of generation of gases that affect the ozone layer, on the other hand, from the social field EGP has generated jobs for people in the area, this in order to take advantage of the national workforce and provide academic training that allows the population a better quality of life while the project is in execution [24]. Some of the features that stand out most is the implementation of bifacial panels, although taking into account the area in which it is installed, it is then limited to the generation capacity since for this area it is expected that energy can be generated 2 or 3 times more than what they currently produce.

4) *Argentina*: In the regions of Neuquén to Jujuy, there is the largest solar irradiation that Argentina has, which is why the importance of the implementation of this type of energy generation sources will yield a new horizon at the economic and energy level for the country. In recent years, the creation of photovoltaic plants throughout the country has been

booming, which is why in the world ranking of energy generation it is ranked 42nd, largely supporting the national interconnected system with clean energies [14].

Some economic analysts indicate that the country would save around 300 million fuels if 1000 MW of renewable energy will be implemented, for this reason it is very important to recognize what characteristics one of the large solar parks that the country has can possess. Some of this information is provided below in Table III.

It can recognize then that this type of solar parks has required a large economic investment due to the environmental impacts they generate in the area, these impacts are of a social and ecological nature, therefore it was necessary that some generating companies will merge as is the case of altiplano 200 in order to maximize their profits by grouping each of the parks that had been created in the area [14]. There is an enormous energy potential of the country, but it is necessary that external agents invest this resource, countries such as China in the last decade have generated agreements and negotiations between energy peers of the country in order to establish some new solar plants that maximize the renewable energy generation of Argentina.

#### B. Comparison between Selected Projects

When entering each of the characteristics of the largest photovoltaic solar generation parks that Brazil, Mexico, Chile and Argentina have, it was able to show what strengths and weaknesses are presented in each of these projects, the ones that stand out below will be mentioned.

##### 1) Strengths

- Implementation of state-of-the-art technology such as bifacial panels.
- Social inclusion policies that allow a good execution of these projects in the area.
- Capacity of generation to do with it is in the national energy system.
- Weaknesses.
- Effects on the flora and fauna of some areas.
- Excessive water consumption in the construction and execution of these projects.
- Damage to the ground in the medium and long term.

Although it is seen that large companies that develop this type of projects in Latin America, such as Enel, have multiple mechanisms to mitigate many of the weaknesses mentioned above, but unfortunately in some areas where ecology thrives, environmental problems are occurring whose consequences are irreversible. That is why a call is made for this type of policies to be more optimized so that the damage to the environment when these projects are developed and executed have been mitigated to the maximum and not only focus on the reduction of greenhouse gases since this is not the only environmental damage that can be generated.

Loss of energy in solar panels is usually due to two factors, which correspond to the temperature and dirt that are detrimental in photovoltaic installations; where between 15% and 25% of the energy generated can be lost by photovoltaic modules. When designing a photovoltaic installation, consumption must be taken into account and also estimate the number of panels to be installed for the energy demand, in order to correct possible increases or decreases in electrical energy. To mitigate the problems in terms of dirt that occurs in the panels due to climatic factors, a possible solution is to manually clean said modules for the conservation of the photovoltaic plant.

VI. CHALLENGES AT THE LEVEL OF RESEARCH AND EXECUTION OF PHOTOVOLTAIC PROJECTS IN LATIN AMERICA

The gap opens so that researchers and education centers related to renewable energy generation can inquire about the problems that currently afflict this type of energy systems such as:

- Improve the efficiency of solar panels.
- Regulate the stability of the system.
- Mitigate the impacts on generation due to pollution.
- Portable implementation of these systems at low cost.
- Use of solar cells in public and private means of transport.

Among other projects or mechanisms that optimize the use of these systems.

The technologies that implement photovoltaic systems are relatively recent date back about four decades, which is why there are multiple problems derived from this type of technology thus preventing the increase of its efficiency gradually, the reasons affect the science of these systems are shown in Table IV.

Taking into account the aforementioned affectations, it is necessary to investigate technologies or mechanisms ken to minimize this type of losses to the photovoltaic system that is why some of these are mentioned in order to provide approach tools for future research related to this topic.

TABLE IV. EFFECTS ON SOLAR ENERGY SYSTEMS

Problems	Description
Climate impacts	The parameters that affect are the temperature of the surface of the panel and its corresponding radiation, climatic factors, wind and humidity of the area
Environmental impacts	Dust pollution in solar panels, spaces where shadows can be generated, soil degradation, etc.
Characteristic features	Damage to the structure that holds the solar modules, failure of solar cells, elements tearing, etc.
System losses	Losses in conductors, inverters, material with which the cells are manufactured, environmental contamination and distortion of solar radiation. On average, losses are estimated at 25% of the estimated generation.

A. Optimization Mechanisms

1) Control of system stability failures: Based on the great boom in photovoltaic generation worldwide, it can appreciate some challenges or problems arising from this implementation such as voltage stability, power oscillations and frequency regulations. The energy oscillations of the system They are a situation that occurs mostly in areas where its energy flow is directly derived from photovoltaic systems, which is why for this work it is imperative to establish parameters to control the oscillations thus guaranteeing the reliability of the system. [22].

The proposed strategy to mitigate this impact is the implementation of a new control mode that applies to the inverter that is in the photovoltaic network and in the storage system at the moment where the energy oscillations that are parallel to the frequency variations and the state of the load of the storage cell are presented, thus imitating the maximum and minimum values of load and discharge [22]. The hysteresis control allows avoiding the frequent change of the working modes, on the other hand, initially the battery has the role of moderating the energy that comes from the outside, a reactive power controller is also attached to this system, and these mechanisms are used according to the energy level of the battery as shown in Fig. 9.

2) Model for detecting dust accumulation in solar cells: The generation of energy in a photovoltaic system depends directly on the efficiency of its modules or solar cells, this parameter changes according to external factors such as the dirt that can occur in these elements generally in arid areas, as part of the challenges faced by this type of technologies is the mitigation of this type of physical pollution. To detect such a change, it is necessary to make the relevant measurements to a new or recently implemented photovoltaic system [23].

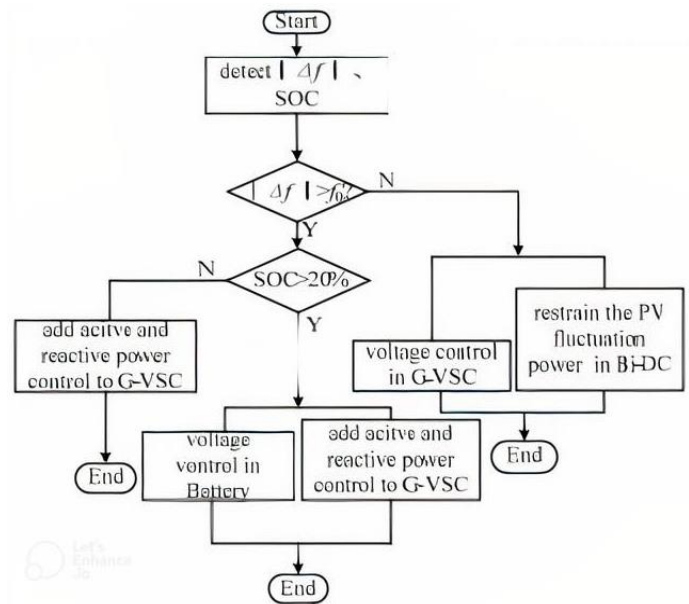


Fig. 9. Gram flow of the control coordination strategy [22].

As an object of this detection model, corresponding measurements of radiance, accumulated power, and temperature in a time interval in a desert of Saudi Arabia are implemented to establish the prediction model. The results of this analysis were that during a year in the months where more sandstorms occurred that is about half of the year, a loss of up to 20% of energy reception was created, thus recognizing the importance of implementing some technological mechanism that limits dust pollution in those months of the year where more losses are generated.

3) *Hybrid systems*: Starting from the premise that photovoltaic systems can only generate energy during the day, it is necessary to find a mechanism that allows generating energy at night. As a result of this problem, the connection of a wind system that allows generating energy at night has already been implemented so that it can be used 24 hours a day in energy generation [23].

Some of the limitations of the implementation of this type of system are that they can only be developed in isolated areas where power lines are limited. It should also be taken into account that to store this energy derived from a wind turbine, the transformation from AC to DC must be generated in order to regulate the load in order to adjust it to an energy that can receive the storage batteries, which are also implemented by photovoltaic systems.

#### B. New Technologies on the Continent

1) *Two-sided solar panels*: Currently some of the new technologies that are being implemented are the use of bifacial panels which allow the use of the entire area of the solar cells in order to increase the percentage of energy reception.

This type of panels has been used by corporations that implement large solar parks in countries such as Brazil, Chile, Mexico, among others. It is expected that in about 10 years all single-facial solar panels that are used in solar power plants will be updated in order to optimize the performance of this type of generation.

2) *Photovoltaic inverters*: Another of the advances that are being implemented is the use of photovoltaic inverters that control the stability of the system, this in order not to introduce harmonics in the laying by the electricity network in order to avoid losses due to this type of distortions. The strategy is to control the injection of active and reactive power into the system to improve the efficiency of generation and distribution of photovoltaic solar energy.

3) *Implementation in transport*: They are currently conducting studies and have some prototypes of vehicles that have solar panels on their roof that are used to generate electricity in the device. At the moment there are only vehicles with very small panels which means that the energy received is not enough to guarantee the autonomy of mobility in a journey of more than 50 km in hours where the traffic is strenuous. In some countries in Europe, North America and Asia, are very common and the use of solar panels in your transport systems such as buses, electric vehicles, trains or

ships thus allowing the non-dependence on fossil fuels and more now that due to the situation between Russia and Ukraine, the flow of oil and gas to the entire European continent has been limited.

#### VII. CONCLUSIONS

One of the main challenges in the study of the factors that affect the viability of photovoltaic projects is the lack of information on the different parameters such as geographic location, orientation of the installation, the amount of available sunlight, the level of solar irradiation and the level of energy captured. These limitations are important when evaluating the performance of a photovoltaic project.

In the future, photovoltaic projects are expected to be further developed in Central and South America feeding the electrical grid with clean and renewable energy. Additionally, projects can provide local employment and support economic development. These photovoltaic projects are gaining ground and are becoming a priority. This is due to the growing concern about climate change, as well as the intention to reduce imports of fossil fuels. Therefore, when analyzing the data, an overview of the situation in each country is given. Among the most important factors to consider were: the initial capital required, the installation costs, the geographical location and the cost recovery time, for which it is necessary to generate new knowledge and technologies that carry both the collection efficiency energy, either by improving the composition of the materials of each solar panel, adopting more intelligent economic and regulatory strategies, and becoming aware of the limitations that each country presents.

This article offers an overview of the main aspects that should be considered when evaluating a solar project. It can become a reference for students and professionals in the electricity sector, as it offers a concise and complete overview of the main factors to take into account when evaluating the feasibility of a photovoltaic project. This has led to a greater understanding and use of resources, as well as greater efficiency in the decision-making process.

It is recommended that for future work a study be taken into account that considers the effects that weather conditions and materials may have on the efficiency of the panels, since this can significantly affect performance. Therefore, for a complete evaluation of the viability of photovoltaic projects, this impact must be considered, analyzing how the components interact with each of the climatic conditions, emphasizing considering the resources that represent the best cost-benefit ratio and that can better benefit Latin American countries.

Taking steps to reduce the costs of photovoltaics is essential to ensure that the use of renewable energy is as economical as possible. A clear solution to reduce the costs of electricity is the improvement of solar panels, attacking this lack in energy capture by favoring the reduction of the costs of photovoltaic energy. Measures such as increasing research on solar panels, reducing production costs, and improving design must be taken to optimize performance and lifetime. Research and development in solar panels must be a priority to reduce production costs. This must include the development of new



technologies and materials to optimize efficiency. Given that this type of energy once put into operation is a solution in terms of greenhouse gas emissions. Since they do not produce them and also provides a great tool for governments to meet their sustainable energy objectives.

#### REFERENCES

- [1] G. Petrone, C. Ramos y G. Spagnuolo, Photovoltaic sources modeling, Chichester: Wiley, 2017.
- [2] A. Guisández y M. Santos, «Modelling and experimental validation of aging factors of photovoltaic solar cells,» IEEE Lat. Am. Trans, vol. 19, nº 8, pp. 1270-1277, 2021.
- [3] Z. Z. \*. a. M. Carbajales-Dale, «Assessing the photovoltaic technology landscape: efficiency and energy return on investment (EROI),» Energy & Environmental Science, vol. 11, nº 3, pp. 603-608, 2018.
- [4] D. H. Mojoner, A. Rios Villacorta y J. Luyo Kuong, «Impact Assessment of Net Metering for Residential,» INTERNATIONAL JOURNAL of RENEWABLE ENERGY RESEARCH, vol. 8, nº 3, p. 10, 2018.
- [5] L. A. Becerra Pérez, R. R. Gonzalez Diaz y A. C. Villegas Gutierrez, «Photovoltaic solar energy, cost benefit analysis of project in Mexico,» Revista Internacional de Desarrollo regional Sustentable, vol. 5, nº 2, p. 24, 2022.
- [6] Enel, «Enel Green Power Chile comienza la construcción de la planta solar Domeyko en la región de Antofagasta,» Enel Green Power, 2020. [En línea]. Available: <https://www.enelgreenpower.com/es/medios/press/2020/10/enel-green-power-chile-comienza-la-construccion-de-la-planta-solar-domeyko-en-la-region-de-antofagasta>. [Último acceso: 12 09 2022].
- [7] Enel, «Parque solar Ituverava, Brasil,» Enel Green Power, 2021. [En línea]. Available: <https://www.enelgreenpower.com/es/proyectos/operativos/parque-solar-ituverava>. [Último acceso: 12 09 2022].
- [8] Enel, «Nova Olinda,» Enel Green Power, 2017. [En línea]. Available: <https://www.enelgreenpower.com/es/medios/news/2017/2/nova-olinda-llega-a-la-vida-instalado-el-panel-solar>. [Último acceso: 12 09 2022].
- [9] Enel, «Central solar Villanueva,» Enel Power Group, 2018. [En línea]. Available: <https://www.enelgreenpower.com/es/medios/news/2018/3/enel-green-power-mexico-inaugura-villanueva-la-planta-solar-fotovoltaica-mas-grande-de-las-americas>. [Último acceso: 12 09 2022].
- [10] Enel, «Instalado en Magdalena II el primer panel fotovoltaico bifacial,» Enel Green Power, Abril 2019. [En línea]. Available: <https://www.enelgreenpower.com/es/medios/news/2019/04/se-instalo-primer-panel-fotovoltaico-bifacial-magdalena-II>. [Último acceso: 12 09 2022].
- [11] Irena, «Estadísticas de energía renovable,» Abu Dhabi, 2022.
- [12] Enel, «Planta solar Sol de Lila,» Enel Green Power, 2020. [En línea]. Available: <https://www.enelgreenpower.com/es/medios/press/2020/11/en-el-green-power-chile-comienza-la-construccion-del-proyecto-fotovoltaico-sol-de-lila>. [Último acceso: 12 09 2022].
- [13] Jemse, «Solar Cauchari,» Jemse, 2020. [En línea]. Available: <https://jemse.gob.ar/2020/12/04/solar-cauchari-se-puso-en-marcha-el-parque-mas-grande-de-la-region/>. [Último acceso: 12 09 2022].
- [14] Ámbito, «Parque Solar Fotovoltaico Altiplano 200,» 2021. [En línea]. Available: <https://www.ambito.com/energia/salta/inauguro-uno-los-parques-solar-mas-grandes-del-pais-n5336106>. [Último acceso: 12 09 2022].
- [15] C. R. S. Júnior, G. C. Luiz, G. A. S. Barbosa y L. G. C. Sousa, «Aerosol Influence on Photovoltaic Solar Energy Generation for the State of Goiás,» de IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS), Santiago, Chile, 2020.
- [16] A. Calazans, M. Kelly, G. Chaudhry y M. K. Siddiki, «Economic analysis of a photovoltaic system connected to the grid in Recife, Brazil,» de IEEE 42nd Photovoltaic Specialist Conference (PVSC), New Orleans, LA, USA, 2015.
- [17] A. E. Alves, D. F. d. Santos y J. V. d. Silva, «Performance of photovoltaic systems integrated in buildings (BIPV): a case study on campus Itaperuna of the Fluminense Federal Institute Brazil,» de IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America), Gramado, Brazil, 2019.
- [18] J. L. d. S. Silva, K. B. d. Melo, T. S. Costa, G. M. V. Machado, H. S. Moreira y M. Gradell, «Impact of Bifacial Modules on the Inverter Clipping in Distributed Generation Photovoltaic Systems in Brazil,» de Brazilian Power Electronics Conference (COBEP), João Pessoa, Brazil, 2021.
- [19] D. V. d. S. Stilpen y V. Cheng, «Solar photovoltaics in Brazil: A promising renewable energy market,» de 3rd International Renewable and Sustainable Energy Conference (IRSEC), Marrakech, Morocco, 2019.
- [20] Neto, R. D. S. N, Lopes, F. M. de Oliveira Carneiro, T. L y Junior, B. F. S., «ANÁLISE COMPARATIVA ENTRE A IMPLANTAÇÃO, OPERAÇÃO E COMERCIALIZAÇÃO DE ENERGIA ELÉTRICA DE EMPREENDIMENTOS SOLARES E TÉRMICOS,» ANAIS. VII Congresso Brasileiro de Energia Solar-CBEN, vol. 1, pp. 1-8, 2020.
- [21] X. Zhang, I. yang y X. Zhu, «Integrated control of photovoltaic-energy storage system for power oscillation damping enhancement,» IEEE 8th International Power Electronics and Motion Control Conference (IPEMC-ECCE Asia), pp. 1571-1575, 2016.
- [22] M. A. Kumar, M. Alaraj, I. Rizwan, A. a. M y Jamil, «Development of Novel Model for the Assessment of Dust Accumulation on Solar PV Modules,» IEEE Journal of Photovoltaics, 2022.
- [23] H. HassanzadehFard, S. M. Hakimi y A. Hasankhan, «Hybrid Energy Storage Systems for Optimal Planning of the Heat and Electricity Incorporated Networks,» Coordinated Operation and Planning of Modern Heat and Electricity Incorporated Networks, pp. 481-508, 2023.
- [24] A. Tartaglia, Tecnología fotovoltaica con paneles bifaciales: ¿vuelve para quedarse?, Cartagena: UNIVERSIDAD POLITÉCNICA DE CARTAGENA, 2019.
- [25] T. Diaz y G. Carmona, Instalaciones solares fotovoltaicas, Madrid: Mc Graw Hill, 2010.
- [26] G. S. Ruiz, S. I. Valencia, I. J. Guevara y R. J. López, «Comparative of wind systems vs photovoltaic for the implementation in the electric network of Veracruz Port,» de IEEE International Conference on Engineering Veracruz (ICEV), Boca del Rio, Mexico, 2019.
- [27] Rayan Ribeiro Silva, G, Inserção da energia solar e eólica no Brasil e nas grandes potências mundiais., Anápolis: Univerisdade Estadual de Goiás, 2022.

# Ransomware: Analysis of Encrypted Files

Houria MADANI<sup>1</sup>, Noura OUERDI<sup>2</sup>, Abdelmalek Azizi<sup>3</sup>  
Faculty of sciences, Mohammed First University, Oujda, Morocco<sup>1,2,3</sup>

**Abstract**—Ransomware is a type of malware that damage the system by encrypting all the files existing in the computer. To get access, the victim has to pay a ransom to get a key to decrypt his data. When the virus is running in machine, the user cannot stop it on the first try, so he may lose his entire files. One of the goals of this work is to detect ransomware based on encrypted files in real time and to minimize the cost of losing files. We will try to do an analysis of a received file (without opening it and seeing its contents). This scanning action can prevent a ransomware from spreading in the system. Most Ransomware files are sent in “.exe” format, but in this work, we will try to use other file formats that can accept malware, for example, .doc or .docx, .xls or .xlsx, .ppt or .pptx, .jpg, etc. In fact, an attacker can focus only on the files that contain useful data. In this paper, we are going to identify the types of files if they are suspicious or normal (without opening them) from their headers. For that first, we are going to analyze each extension separately (.docx, .exe, .pptx, .xlsx, .jpg, etc.) by identifying their headers and signatures. Then we will take several files with different extensions to analyze them by doing a program who detect if a file is benign or suspicious.

**Keywords**—Ransomware; encrypted files; signature; file format; static analysis

## I. INTRODUCTION

In recent years, ransomware attacks continue to explode exponentially around the world; the cost keeps falling and exploit different sectors.

Researchers and cybersecurity specialists are still looking for a solution to detect this attack and even to slow down its growth in order to find an effective and reliable solution. We see many solutions, but not 100% sure, because hackers are always attentive and updated with the new technologies, they use more sophisticated techniques to follow the evolution and bypassing the protection techniques.

This study focuses on the examination of the behavior and method in which ransomware encrypts files. Ransomware can infiltrate a device in various formats like .exe, .docx, .ppt, etc. A user may open a .docx file without realizing it is an unsafe file that contains metadata that can damage their computer. Therefore, we aim to analyze the files (without opening them) before and after ransomware encryption, in order to distinguish between a typical file and a suspicious one.

In this paper, we will make a study on files to differentiate between a normal file and a suspicious one. For that in Section II, we will approach some "state of the art" concerning the study of files to give you an idea of the current research on this subject. In Section III, we will see our objectives and working methodology to identify and detect a normal file from another suspect one. We will discuss the results that we have

had in Section IV. At the end, we sum up with a conclusion and some perspectives.

## II. STATE OF THE ART

As you know, attackers are very inventive when they want to target a victim and we find, often, that emails are the trickiest (more than 90%) way [1] for them to create a link between the attacker and the target. Fig. 1 explains how ransomware attacks your machine:

Ransomware detection techniques [2]–[5] are becoming more and more competitive, and each researcher has his own method and technique. If we take the detection of ransomware or malware in general, using file headers, several researchers work focus on a single file extension like PE (Portable Executable) files [6]–[8], but there is not enough research on the detection of ransomware using the headers of different extensions.

The authors in [9] proposed a new classification model based on machine learning techniques to detect and classify malicious and benign PE files based on their headers information. The experimental results proved that the Random Forest algorithm yields a higher accuracy (99.68%) compared to other algorithms. The tests were performed on 211,067 malware samples obtained from the VirusShare database [10]. Manavi and Hamzeh [11] presented a method for detecting ransomware using the PE header. They used a Convolutional neural network (CNN) to identify ransomware by converting the header bytes into 32\*32 pixel images. The use of a header is advantageous, but transforming it into an image would necessitate the use of a network with additional layers in order to extract its features.

To detect ransomware, the authors [12] used a static method. They proposed a method that is based on the bytes extracted from the header of the executable file using LSTM network to build the detection model.

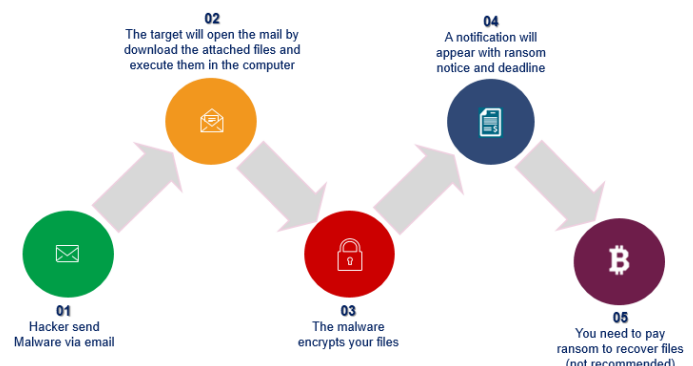


Fig. 1. Ransomware attack phases.

\*Corresponding Author.

The modification of the file header changes its structure. Therefore, they did the extraction of the executable file headers, then they processed the byte sequence that builds the file header with LSTM network, and they separated the ransomware samples, from the benign samples to form the template. With this technique, they managed to detect ransomware with 93.25% accuracy without running the program.

Subedi et al. [13] employed data mining techniques to recognize and detect ransomware families using both static and dynamic analysis at three different levels: assembly, function calls and library. They also created an analytical tool that uses reverse engineering to create signatures for identifying ransomware families. Arabo et al. [14] proposed a dynamic analysis approach to gather ransomware API properties, which is then utilized to test 9 Machine Learning classifiers and a neural network. The goal of this research is to understand the link between a process's behavior and its nature, to detect if it is a ransomware or not. With a detection rate of 75.01%, Random Forest surpasses other classifiers. The benefit of this technique is that it does not require a signature database, but rather a collection of ransomware and non-ransomware data. The detection rate of the classifiers may be better by improving the dataset.

Before encrypted files were moved to a backup disk, Lee et al. [15] utilized machine learning techniques to detect and classify infected files. The training step was implemented at the backup system according to their recommendation. It identified files from various users and file types, as well as determining file entropy thresholds. These thresholds were transmitted to client hosts in order to decide whether a new version of the file was encrypted or not. The authors in [16] suggest a two-stage mixed ransomware detection approach using Markov model with the Random Forest technique to detect ransomware. Random Forest has the best detection rate of 97.3%.

The paper [17] emphasizes the capabilities of behavior-based detection mechanisms to identify crypto ransomware, demonstrating the limitations of signature-based detection approaches. In [18], Nieuwenhuizen proposed a ransomware detection scheme using behavior analysis and machine learning. Although the specific features were not revealed, their created feature set included properties such as payload persistence, anti-system restoration, stealth methods, environment mapping, network traffic, and privilege elevation that were extracted from the behavior of a malicious set up. Author employed the support vector machine (SVM) method as the classification technique in addition to the behavioral features related with data transformation behavior, such as huge file encryption.

The effect of certain ransomware families on the Windows platform is demonstrated and analyzed by Mohammad [19]. He deduces that most families of ransomware behave in a similar way when it comes to affect file system and registry entities. Furthermore, all types of ransomware generate files in the Windows system files and rename other files. To do the experiments, the author used Windows 7, Oracle VirtualBox VM, Cuckoo sandbox, and Virtual windows 10. The author concludes that monitoring system file and registry activities

can protect against ransomware. He also mentions that Windows 10 is more effective than Windows 7 regarding malware. The best method to follow as a recommendation is to regularly back up company or individual data.

### III. METHODOLOGY

As mentioned at the beginning, our goal is to detect whether a file is suspicious or not (regardless of its content), from its header which will be identified from its extension. This leads us to detect ransomware from encrypted files in real time.

It is well known that each extension has a fixed header according to the standards. If the header differs from the standard state, we deduce that it is suspicious.

To achieve our goal, we took several files with different extensions, if we take an extension, for example “.docx”, and we open some files with the same extension using the Hexadecimal editor (Hex Editor Neo [20]), we found that they have the same signature, also called "Magic number".

Namely, each file extension (.docx, .pptx, .xls, .exe, .dll, .jpg, etc.) has a fixed and specific "Magic number".

Table I shows some files with different extensions and their signatures or Magic number, for clear and normal files.

According to a deep study on Microsoft Office files, we notice that their signature is different, the "x" added at the end made many differences. If we take the extensions .doc and .docx (the same thing for .xls, .ppt/.xlsx, .pptx), the differences are seen in Table II.

TABLE I. SIGNATURE OF FILES WITH DIFFERENT EXTENSIONS

File extension	Hex signature	Size	ASCII Signature
.doc , .ppt , .xls	<b>D0 CF 11 E0 A1 B1 1A E1</b>	8 Bytes	ĐĬ.ą±.á
.docx , .pptx , .xlsx	<b>50 4B 03 04 14 00 06 00</b>	8 Bytes	PK.....
.pdf	<b>25 50 44 46</b>	4 Bytes	%PDF-
.png	<b>89 50 4E 47 0D 0A 1A 0A</b>	8 Bytes	.PNG....
.jpg	<b>FF D8 FF E0</b>	4 Bytes	ÿØÿà
.dll	<b>4D 5A 90 00</b>	4 Bytes	MZ
.exe	<b>4D 5A</b>	2 Bytes	

TABLE II. DIFFERENCE BETWEEN .DOC AND .DOCX

	DOC	DOCX
<b>Version</b>	Came in when the Microsoft Word was 1 <sup>st</sup> delivered and was utilized until 2003 variant of "Word".	Came in with Word 2007 and has been the default extension from then for all new Word versions.
<b>Storage</b>	A DOC is saved in a binary file that contains all the related formatting and relevant informations.	A DOCX file is actually a zip file with all XML files associated with the document.
<b>File size</b>	The DOC format has a <i>greater</i> size than the DOCX format.	The DOCX format has a <i>smaller</i> size than the DOC format.

To analyze files, we have taken some files (.docx, .doc, .pdf, .exe, .png) as an example, and see their structures (this study of files is very difficult to make because there is a lack of information on the structure of the files, for example, for the header "how many bytes occupied", for the contents "where does it begin?", etc.):

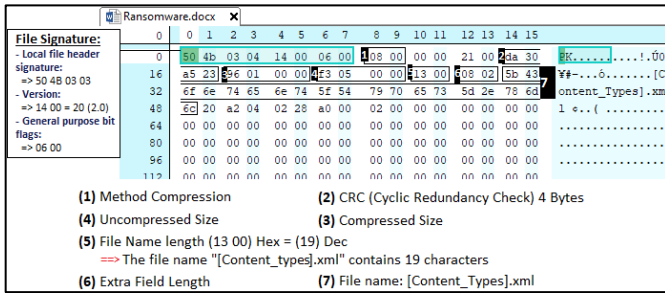


Fig. 2. ".Docx" File signature.

From the article [21], the header is always at the beginning of the file and is exactly 512 bytes in length. Fig. 2 and Fig. 3 show you some information about the header of the ".docx" and ".doc" file, respectively.

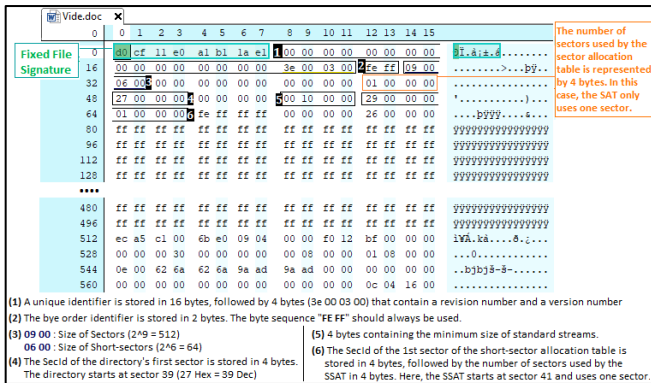


Fig. 3. ".doc" File signature (empty file).

Fig. 4 and 5 show you the structure and header of the ".pdf" and ".exe" files, respectively.

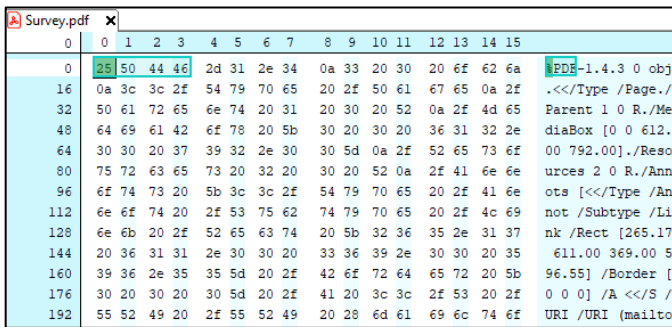


Fig. 4. ".PDF" File Signature.

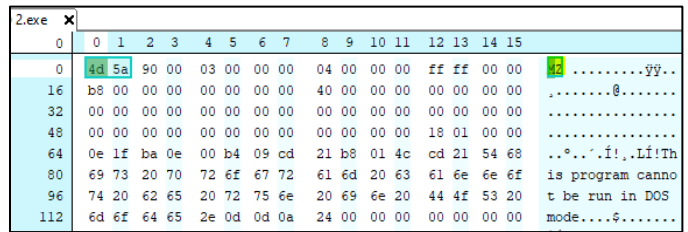


Fig. 5. ".exe" file signature.

#### IV. RESULTS AND DISCUSSION

We took a corpus[22] that contains a large number of files with different extensions, and I encrypted them with a python program, adding to its files an '.enc' extension to make the difference between a clear file and an encrypted file. As an example, I took four files for each different extension (.doc, .docx, .ppt, .pdf); we got the following result:

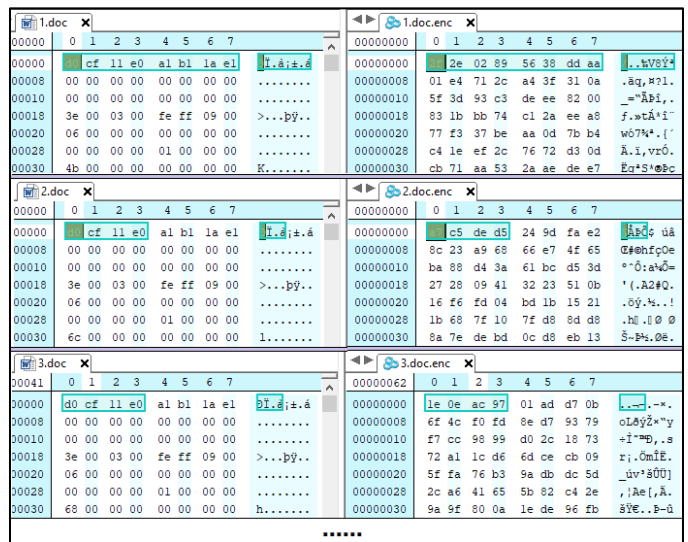


Fig. 6. The difference between encrypted and clear file headers [.doc].

For the files ".doc" (in Fig. 6), those on the left are clear files, their header should be normal [D0 CF 11 E0 A1 B1 1A E1]. While on the right, you see that there is an extension added at the end ".doc.enc", this means that they are encrypted files (the encrypted file of each clear file, e.g. "1.doc.enc" is the encrypted file of "1.doc"), and even their header is different. What is relevant is that each encrypted file has a different header from the other file, we have [2F 2E 02 89 56 38 DD AA], [A7 C5 DE D5 24 9D FA E2], etc.

The same thing for the files ".docx" (in Fig. 7), those on the left are clear files, their header should be [50 4B 03 04 14 00 06 00], while on the right you see the encrypted files and even their header is different. We have [A1 D1 05 C2 6C 8B 1D 19], [EF 34 21 9F FE 78 65 FC], etc.

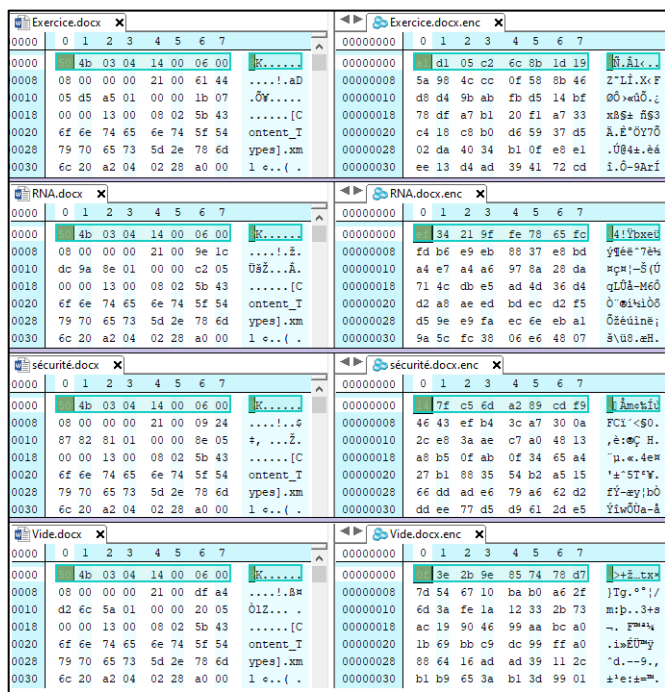


Fig. 7. The difference between encrypted and clear file headers [.DOCX].

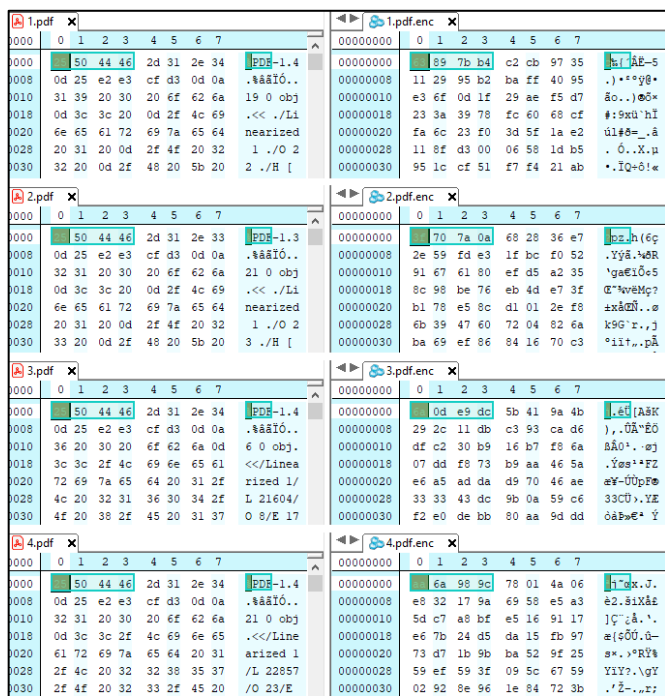


Fig. 8. The difference between encrypted and clear file headers [.PDF].

As you can see in Fig. 6, Fig. 7 and Fig. 8, the signature of a clear file and its encrypted is not the same; the case is the same for the other extensions. We also notice that the signature is fixed for any clear file (with different extension), but for encrypted files, it is not fixed and differs between each file.

The program is done by Python language to make this study and detect if a file is suspicious or normal from its signature.

Each extension has a "Magic Byte". We instantiated our dataset by creating a dictionary with the file extension as a key and its "Magic byte" as value, and then we analyze the file. If the file does not contain the corresponding signature, i.e. it has a different header than the one presented in our dataset; we deduce that it is a suspect file. We have also dealt with the case of a file without extension, if we give it to our program, it analyzes the header and if it does not find the corresponding signature, it sends us back that it's a suspicious file, otherwise, if everything is normal the result is: "This is a benign file, its extension is: ... ". Fig. 9 shows the result of a file without extension that is benign.

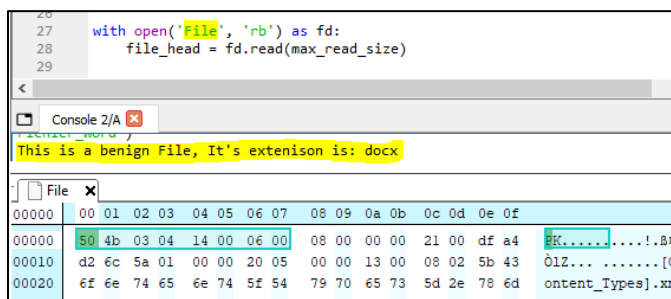


Fig. 9. Analysis of a file without extension.

Fig. 10 shows you that we have performed an analysis for several files with different extensions.

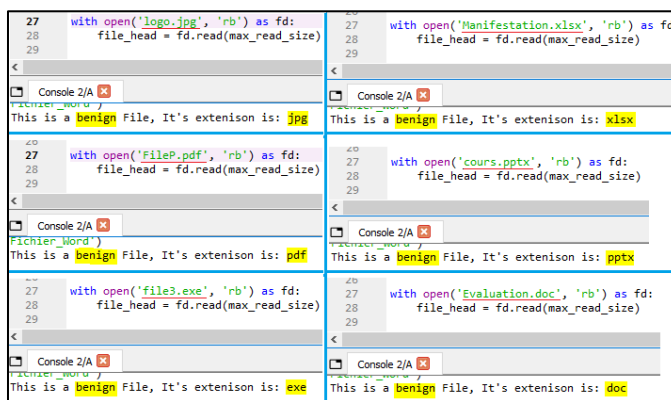


Fig. 10. Analysis of several normal files with different extensions.

If we take the example in Fig. 11, you can see that the result is "this is a suspicious file", even though the file has the extension ".doc". In effect, sometimes attackers send files that look normal with a legal extension, while the file is infected by the ransomware, so as you can see, our program perfectly analyzes the header of the given file identifying its signature, and it found that its signature does not match to the normal signatures.

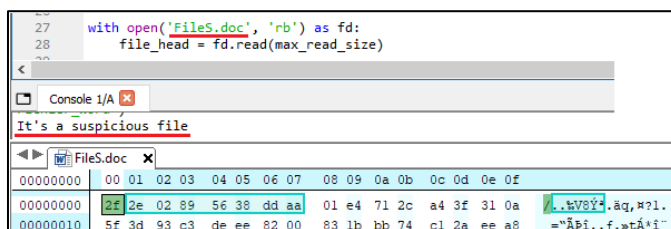


Fig. 11. Analysis of a suspicious file.

We have tested our program on files encrypted by Ransomware with ".lmas", as you can see in Fig. 12, we have taken as an example the file "formation.xlsx.lmas", the ".lmas" extension is added after ".xlsx" extension. We got the followed result:

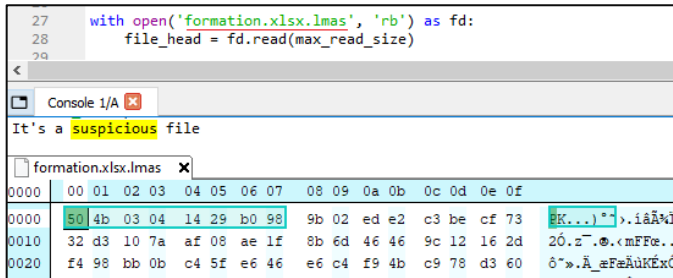


Fig. 12. Analysis of a file encrypted by Ransomware.

As you can see, Ransomware infects the file «formation.xlsx», it is encrypted and the attacker has added the extension ".lmas" to the file.

We know that ".xlsx" has a fixed signature (see Table I); in the Fig. 12, we can see that the first 4 bytes are similar to the first 4 bytes of the normal xlsx file (50 4B 03 04), but the difference is in the next 4 bytes. Therefore, our program was able to detect that this file is encrypted by ransomware so it is a suspicious file without opening it.

## V. CONCLUSION

In this work, we have made a Python program that allows to detect a suspicious file from another normal one, we started by studying the header of files of different extensions separately, later we extracted the header of each file and compared the headers of a normal file with another encrypted one. With this study, we could deduce that a normal file has a fixed and unchangeable extension, once it is changed the file is suspicious.

In the upcoming work, we will conduct a dynamic analysis by executing ransomware files in a simulated environment. This will allow us to extract ransomware encrypted files and analyze them in order to develop and implement our own neural network. This network will be trained to identify ransomware files by first learning the characteristics extracted from the ransomware encrypted files, and then using that knowledge to detect ransomware when a "vulnerable" file is downloaded onto a victim's device.

## REFERENCES

[1] « Cyberattacks 2021: Statistics From the Last Year », Spanning, 18 janvier 2022. <https://spanning.com/blog/cyberattacks-2021-phishing-ransomware-data-breach-statistics/> (consulté le 7 mars 2022).

[2] S. Kok, A. Abdullah, N. Jhanjhi, et M. Supramaniam, « Ransomware, threat and detection techniques: A review », *Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no 2, p. 136, 2019.

[3] C. V. Bijitha, R. Sukumaran, et H. V. Nath, « A Survey on Ransomware Detection Techniques », in *Secure Knowledge Management In Artificial Intelligence Era*, vol. 1186, S. K. Sahay, N. Goel, V. Patil, et M. Jadliwala, Éd. Singapore: Springer Singapore, 2020, p. 55-68. doi: 10.1007/978-981-15-3817-9\_4.

[4] C. Beaman, A. Barkworth, T. D. Akande, S. Hakak, et M. K. Khan, « Ransomware: Recent advances, analysis, challenges and future research directions », *Computers & Security*, vol. 111, p. 102490, déc. 2021, doi: 10.1016/j.cose.2021.102490.

[5] S. I. Bae, G. B. Lee, et E. G. Im, « Ransomware detection using machine learning algorithms », *Concurrency Computat Pract Exper*, vol. 32, no 18, sept. 2020, doi: 10.1002/cpe.5422.

[6] S. Poudyal, K. D. Gupta, et S. Sen, « PEFile Analysis: A Static Approach To Ransomware Analysis », *The International Journal of Forensic Computer Science*, vol. 14, p. 34-39, oct. 2019, doi: 10.5769/IJ201901004.

[7] T. Rezaei et A. Hamze, « An Efficient Approach For Malware Detection Using PE Header Specifications », in *2020 6th International Conference on Web Research (ICWR)*, Tehran, Iran, avr. 2020, p. 234-239. doi: 10.1109/ICWR49608.2020.9122312.

[8] V. Verma, S. K. Muttoo, et V. B. Singh, « Detection of Malign and Benign PE Files Using Texture Analysis », in *Information Systems Security*, Cham, 2020, p. 253-266. doi: 10.1007/978-3-030-65610-2\_16.

[9] I. Abdessadki et S. Lazaar, « A New Classification Based Model for Malicious PE Files Detection », *International Journal of Computer Network and Information Security*, vol. 11, p. 1-9, juin 2019, doi: 10.5815/ijcnis.2019.06.01.

[10] « VirusShare.com ». <https://virusshare.com/> (consulté le 25 février 2022).

[11] F. Manavi et A. Hamzeh, « A New Method for Ransomware Detection Based on PE Header Using Convolutional Neural Networks », in *2020 17th International ISC Conference on Information Security and Cryptology (ISCISC)*, Tehran, Iran, sept. 2020, p. 82-87. doi: 10.1109/ISCISC51277.2020.9261903.

[12] F. Manavi et A. Hamzeh, « Static Detection of Ransomware Using LSTM Network and PE Header », in *2021 26th International Computer Conference, Computer Society of Iran (CSICC)*, Tehran, Iran, mars 2021, p. 1-5. doi: 10.1109/CSICC52343.2021.9420580.

[13] K. P. Subedi, D. R. Budhathoki, et D. Dasgupta, « Forensic Analysis of Ransomware Families Using Static and Dynamic Analysis », in *2018 IEEE Security and Privacy Workshops (SPW)*, San Francisco, CA, mai 2018, p. 180-185. doi: 10.1109/SPW.2018.00033.

[14] A. Arabo, R. Dijoux, T. Poulain, et G. Chevalier, « Detecting Ransomware Using Process Behavior Analysis », *Procedia Computer Science*, vol. 168, p. 289-296, 2020, doi: 10.1016/j.procs.2020.02.249.

[15] K. Lee, S.-Y. Lee, et K. Yim, « Machine Learning Based File Entropy Analysis for Ransomware Detection in Backup Systems », *IEEE Access*, vol. 7, p. 110205-110215, 2019, doi: 10.1109/ACCESS.2019.2931136.

[16] J. Hwang, J. Kim, S. Lee, et K. Kim, « Two-Stage Ransomware Detection Using Dynamic Analysis and Machine Learning Techniques », *Wireless Pers Commun*, vol. 112, no 4, p. 2597-2609, juin 2020, doi: 10.1007/s11277-020-07166-9.

[17] P. S. Goyal, A. Kakkar, G. Vinod, et G. Joseph, « Crypto-ransomware detection using behavioural analysis », in *Reliability, Safety and Hazard Assessment for Risk-Based Technologies*, Springer, 2020, p. 239-251.

[18] D. Nieuwenhuizen, « A behavioural-based approach to ransomware detection », *Whitepaper. MWR Labs Whitepaper (2017)*, p. 20.

[19] A. Mohammad, *Analysis of Ransomware on Windows platform*, *International Journal of Computer Science and Network Security* 20.6 (2020): 21-27. 2020. doi: 10.13140/RG.2.2.11150.59202.

[20] « Free Hex Editor: Fastest Binary File Editing Software. Freeware. Windows ». <https://www.hhdsoftware.com/free-hex-editor> (consulté le 25 février 2022).

[21] D. Rentz, « The Microsoft Compound Document File Format », [Internet]. Available: <http://www.openoffice.org.zaxiproxy.com>, p. 25, août 2007.

[22] « Digital Corpora Downloads: corpora/files/govdocs1/threads/ ». <https://downloads.digitalcorpora.org/corpora/files/govdocs1/threads/> (consulté le 7 mars 2022).

# MMZ: A Study on the Implementation of Mathematical Game-based Learning Tool

Nur Syaheera Binti Sulaiman<sup>1</sup>, Hamzah Asyrani Bin Sulaiman<sup>2</sup>, Nor Saradatul Akmar Binti Zulkifli<sup>3</sup>, Tuty Asmawanty Binti Abdul Kadir<sup>4</sup>

Faculty of Computing, University Malaysia Pahang, Pahang, Malaysia<sup>1,3,4</sup>

Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia<sup>2</sup>

**Abstract**— Mathematic has always been one of the hardest subject to be learnt during school. This is the very same issue that student have been facing no matter the level of education that they are in, and this is the reason why Math Maze Zone or also known as MMZ had been developed. MMZ is a mathematical game-based learning tool for primary school students to help them prepare for their final examination for the mathematics subject. The proposed game will consist of mathematical questions and the game design will be focusing on a maze where user will have to search for a way out from the maze while going through the checkpoint within the maze. The checkpoint will consist of mathematical questions and when user answered correctly, they will be able to continue their journey to explore the maze. MMZ focused on 1 chapter only for now which is Chapter 8: Space and Shape. Although the game only consists of one-chapter, primary school students can play the game to enhance their knowledge and to help them to be more engaged with mathematics subject by playing the game. The results will be taken from the perspective of adult who have close relationship with standard six students. Five main sections will also be identified along with MMZ to ensure getting a good result.

**Keywords**—Primary school; educational mathematic game; game components

## I. INTRODUCTION

Having a high level of proficiency in math is vital for the economic success and growth of societies, as math is used in a wide range of fields and industries, including finance, science, engineering, and technology. Developing strong math skills can also open numerous career opportunities and increase an individual's earning potential [1][2]. There are few reasons as to why this is happening in which some students may not be interested in math, which can make it harder for them to learn and retain information. Other than that, Math can be abstract and difficult to understand, especially if students have not developed a strong foundation in math concepts. Moreover, Math also requires practice and repetition to fully understand and master concepts. If students do not have enough opportunities to practice or do not have access to support and resources, it can be harder for them to learn math [3][4].

With pandemic happening and children are unable to go to school for a long period of time, they are relying on the interactive media technology to gain more information and knowledge while staying at home. Not to mention, these children, especially standard 6 students in primary school will still have to take an important exam which will determine

which high school will they further their studies in. One of the methods for them to gain knowledge while learning for their future examination is through watching education videos on YouTube of, they can also play educational games that is related to their school syllabus. Higher availability, flexibility, and affordability of tertiary online education have also proved to have the potential to enable high-quality student learning. [5]

For this research, developing a mathematical interactive game that is suitable and matched the syllabus with primary school student to prepare for their examination will be done. The game will be called MMZ and will consist of mathematical questions. The game design will be focusing on a maze where user will have to search for a way out from the maze while going through the checkpoint within the maze.

The checkpoint will consist of mathematical questions and when user answered correctly, they will be given points. This research will be aimed to develop an educational mathematical game design that will help the primary school student where they can utilize the mathematics questions for their examination. MMZ will also include 5 different components to measure the effectiveness of the game such as the learning comprehension, user-friendliness, process of user learning, creative and critical thinking, and the effectiveness of MMZ towards the user.

There will be 4 sections that will be discussed in this research, which is section II, Background of the research that consist of mathematics examination and existing similar mathematic games. Section III, Methodology which consist of experimental setup and game design. Section IV, Results, and Section V, Conclusion.

## II. BACKGROUND OF RESEARCH

### A. Mathematics Examination

In Malaysia, students must take exams and assessments to move forward in their education. These results help them have a better chance to continue their studies. Although it is not the best way to measure a student's performance, it is currently the only way. For example, primary students must take final exams to have a chance to attend prestigious schools. Even though the UPSR exam is no longer used, students still need to learn math to move on to high school, especially as it relates to physics [6].

In Malaysia and worldwide, Mathematics is a crucial subject in schools. Students must master arithmetic to be

successful in math. Without this skill, they will have difficulty solving real-world problems. Since problem-solving is a vital part of math education, researchers have studied the success of problem-solving lessons [7]. To get into prestigious secondary schools, students need to do well in Mathematics, as it is a crucial subject. According to the National Council of Teachers of Mathematics, math is a widely used language that is important for communicating with people from different cultures and countries [8]. Other than that Mathematics are also a compulsory passing subject in Malaysian public examinations such as entrance examination and SPM.

In 2019, many students in Malaysia received an E grade in mathematics on the UPSR exam, ranking it as the second highest subject for E grades, according to the Ministry of Education Malaysia. English had the highest percentage of E grades, with 23.34% in writing and 14.87% in comprehension with Mathematics as the second subject with a 16.87% E grade percentage [9]. Below is the graphical illustration related to the comparison of percentages in every subject between 2018 and 2019 that can be seen in Fig. 1.

Kod & Nama Mata Pelajaran	Tahun	Peratus						Bilangan Calon
		A	B	C	D	Mencapai (A → D)	E	
011 - Bahasa Melayu-Pemahaman	2019	26.91	30.29	26.66	12.74	96.60	3.40	332,610
	2018	26.99	29.95	25.68	13.28	95.90	4.10	328,939
012 - Bahasa Melayu-Penulisan	2019	29.39	32.27	20.60	13.07	95.33	4.67	332,694
	2018	25.96	31.44	24.10	13.14	94.64	5.36	328,990
013 - Bahasa Inggeris-Pemahaman	2019	11.56	19.68	23.19	30.70	85.13	14.87	332,660
	2018	12.55	19.32	22.33	28.91	83.11	16.89	329,021
014 - Bahasa Inggeris-Penulisan	2019	11.54	8.86	21.04	35.22	76.66	23.34	332,648
	2018	11.18	7.02	20.17	36.18	74.55	25.45	329,024
015+025+035 - Matematik	2019	19.43	16.84	16.63	30.23	83.13	16.87	431,610
	2018	18.22	15.52	16.96	29.80	80.50	19.50	427,126
018+028+038 - Sains	2019	9.76	30.53	35.95	20.33	96.57	3.43	431,635
	2018	10.31	31.25	33.77	20.79	96.12	3.88	427,151
021 & 031 - Bahasa Melayu-Pemahaman	2019	14.77	24.02	24.00	26.90	89.69	10.31	99,059
	2018	17.71	24.35	23.06	24.69	89.81	10.19	98,108
022 & 032 - Bahasa Melayu-Penulisan	2019	25.39	16.04	28.84	17.47	87.74	12.26	99,058
	2018	23.36	17.24	28.41	17.20	86.21	13.79	98,113
023 & 033 - Bahasa Inggeris-Pemahaman	2019	20.69	21.84	23.11	21.42	87.06	12.94	99,086
	2018	20.10	23.56	21.65	23.05	88.36	11.64	98,142
024 & 034 - Bahasa Inggeris-Penulisan	2019	15.51	17.53	23.80	25.40	82.24	17.76	99,088
	2018	14.86	16.45	24.17	24.08	79.56	20.44	98,144
026 - Bahasa Cina-Pemahaman	2019	17.26	31.07	22.83	19.34	90.50	9.50	85,632
	2018	15.85	32.48	21.03	20.03	89.39	10.61	84,876
027 - Bahasa Cina-Penulisan	2019	28.88	23.60	26.34	11.24	90.06	9.94	85,633
	2018	25.93	24.34	28.17	13.54	91.98	8.02	84,876
036 - Bahasa Tamil-Pemahaman	2019	22.81	33.19	19.75	14.24	89.99	10.01	13,448
	2018	19.71	33.42	18.48	16.06	87.67	12.33	13,252
037 - Bahasa Tamil-Penulisan	2019	31.22	34.11	18.69	8.51	92.53	7.47	13,448
	2018	31.54	34.73	18.42	7.70	92.39	7.61	13,255

Fig. 1. Illustration of the comparison of percentage in each subject between 2018 – 2019 [8]

**B. Existing Similar Mathematic Games**

In this part, reviews about several existing games both in Malaysia and other countries that will be discussed along with

other existing games related to mathematics subject. The comparison MMZ with the other three existing games between the games developed by Malaysian and by other countries will also be discussed. There are six existing mathematics games and system that provide a similar kind of function with MMZ but have different kind of approach but are still using the same concept while developing for the game.

It can be observed that many games developed both in Malaysia and other countries share similarities in that they align with school syllabuses, which can aid students in better understanding the topics taught in school. The distinctions can be seen in Tables I and II, which will be presented below. The research conducted by various researchers in Malaysia will also highlight the principles employed, such as Norman's Seven Principles and Mayer's 12 Multimedia Principles in Table I while Table II will provide a comparison of research conducted by researchers from other countries.

TABLE I. COMPARISON OF EXISTING GAMES FROM MALAYSIA

Research Paper	Details	Norman's Seven Principles	Mayer's 12 Multimedia Principles.
Understanding of Number Concepts and Number Operations Through Games in Early Mathematic Education	Focuses on 1 mathematics topic that follows the NPSC, which is comparing the numbers through a game.	1. Visibility 2. Consistency	1. Multimedia Principle 2. Temporal Contiguity 3. Pre-Training
Pedagogical Change in Mathematic Learning: Harnessing the Power of Digital Game-Based Learning	Focuses on grade 1 to 3 mathematics syllabus. Every lesson on the game is different from one another.	1. Visibility 2. Consistency 3. Constraint	1. Multimedia Principle 2. Temporal Contiguity
Learning Math Through Mobile Game for Primary School Students	A mathematics-based game that consist of all the syllabus in each of the grade.	1. Visibility 2. Signifier 3. Consistency 4. Constraint	1. Multimedia Principle 2. Modality 3. Pre-Training
Math Maze Zone – A study on The Implementation of Mathematical Game-Based Learning Tool for Primary School Students From Adult Perspective	A mathematics-based game that consist of few selected chapter/syllabuses	1. Visibility 2. Feedback 3. Affordance 4. Constraint 5. Consistency	1. Coherence Principle 2. Spatial Contiguity Principle 3. Segmenting Principle 4. Multimedia Principle



TABLE II. COMPARISON OF EXISTING GAMES FROM OTHER COUNTRIES

Research Paper	Details	Norman's Seven Principles	Mayer's 12 Multimedia Principles.
Serious Gaming: Environmental Impact Through Math	Focuses on advance mathematics that universities students are learning.	1. Visibility 2. Mapping 3. Consistency	1. Multimedia Principle 2. Pre-Training 3. Redundancy
Space Chain: A Math Game For Training Geometric And Arithmetic Progressions	Assisting teenagers/students in learning about arithmetic and geometric progression in a comfortable and enjoyable manner.	1. Visibility 2. Consistency 3. Mapping 4. Affordance	1. Multimedia Principle 2. Modality 3. Temporal Contiguity
Game Based Learning For Math Learning: Fractions Case Study	Developed for children aged 6 to 10 for teaching fractions, as well as its pedagogic motivation and two preliminary tests with users	1. Visibility 2. Signifier 3. Consistency 4. Feedback	1. Multimedia Principle 2. Modality 3. Segmenting 4. Signaling

### III. METHODOLOGY

In this section, the experimental setup and game design in MMZ will be discussed.

#### A. Experimental Setup

In the game development process, the project can only move forward once all necessary experimental setup, including multimedia and graphic content, has been prepared. The Unity software platform will be utilized to build the application, with C# as the programming language used within Unity. The programming script will be created using Visual Basic software.

For the implementation process, all the planning should be done in action. Materials used in the project such as audio, video and 3D model should be developed properly. The system that has been developed should also be working properly. The application used to develop the project is by using Unity and Visual Basic software should also be downloaded to enable the application to be opened by using a computer. User should also be able to download the application through Google Store or other websites.

Other than that the scripting of the code for the game is working properly need to make sure. In this process, all programming script will be provided and the script will then need to be implemented in the game. Each scene needs to have script implemented in the game development. For each scene of the game, the script is used to instruct the game. For example, C# script is implemented in each of the button in order to help user to navigate to different scene. There is also a UI pop-out to shows what the user achieved after the user

completed a scene in the application. There will also be a script to allow user to play the video and audio in the application.

Next, in the user testing process, the application need to be tested each of its functionality. The user (which is the adults) who played the game will be conducting the user testing by playing the game and giving their feedback related to the game via the questionnaire and surveys in Google Form. This is the process that will help the developer to make some improvement to the game for it to be a better educational tool for the primary students.

Finally, in the evaluation process, the game will be evaluated based on the performance of the game itself whether it is functioning properly or if it is able to solve the problem statement. The game should also be able to achieve all the objectives and goals that have been set up earlier during the planning phase. Feedback will then be used to make some improvement to the game. If the final product does not meet up the expectation, revision will be needed, and the project will have to be re-analyzed so that a better game can be developed. Below is some of the information related to the development of the proposed game, which is the context diagram, use-case diagram, flowchart, storyboard, and the equipment. The gameplay of MMZ can be visualized through the flowchart presented in Fig. 2, as shown below.

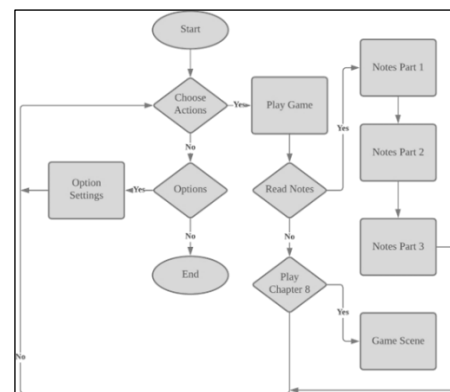


Fig. 2. Flowchart of the gameplay for MMZ

#### B. Game Design

In this section, the game elements, principle of human-computer interactions design and instructional design will be discussed.

##### 1) Game elements

Genre: MMZ is a game that belongs to the maze genre, which involves navigating through a series of twists and turns to reach a specific goal. In this type of game, players are typically faced with a problem that they must solve to progress, such as finding the correct path through a maze or solving a puzzle. By requiring players to think critically and try different approaches until they find a solution, maze games can help learners develop their problem-solving skills and become more strategic thinkers. In the case of MMZ, the goal is for players to navigate through the maze and reach the end, while also solving math problems at each checkpoint along the way. By combining the elements of a maze game with math concepts,

MMZ provides an engaging and challenging way for learners to practice and reinforce their math skills.

**Information:** When players start MMZ, they are provided with some initial information that can help them as they progress through the game. This information includes notes and objectives, which are designed to give players an overview of the game and provide them with the tools they need to succeed. The notes, which can be accessed before the game begins, provide players with important information that can help them understand the concepts being covered and solve the questions they encounter along the way. The objectives, which are also listed in the game, give players a clear sense of what they need to accomplish and allow them to track their progress as they move through the maze. By providing players with this information at the start of the game, MMZ helps to set them up for success and encourages them to stay engaged and motivated as they play.

**Game Objectives:** MMZ is a game that is designed to help learners practice and reinforce their math skills, specifically in shape and space. The objective of the game is for players to navigate through a maze while searching for checkpoints, at which they will encounter questions that test their understanding of math concepts related to shape and space. As they progress through the maze, players will need to use their problem-solving and critical thinking skills to find the correct path to the end, while also looking for an exit from the maze. The inclusion of a chapter on shape and space in the game allows players to focus on a specific topic and deepen their understanding of that concept.

**Quiz:** MMZ includes a quiz feature that allows players to strengthen their knowledge of math concepts. At each checkpoint within the maze, players will be asked a series of questions that test their understanding of the topic being covered. If a player gets a question wrong, they will be given the opportunity to try again, allowing them to loop back and review the material as needed. This iterative process of practicing and reviewing can be an effective way for learners to reinforce their understanding of math concepts and improve their skills over time.

**Game Mechanic:** The game mechanic for MMZ involves searching and exploring a maze while also solving problems at each checkpoint within the maze. As players navigate through the twists and turns of the maze, they will need to use their problem-solving skills to find the correct path to the end. Along the way, they will encounter challenges and obstacles that require them to think critically and apply their math knowledge to proceed. This combination of exploration and problem-solving can be an engaging and challenging way for learners to practice and reinforce their math skills.

**Feedback:** MMZ is a game that is designed to help learners practice and reinforce their math skills. As players progress through the game, they will receive a score or grade that reflects their performance. This can be a useful way for learners to track their progress and understand their strengths and weaknesses. By receiving feedback on their performance, learners can identify areas where they need to improve and focus their efforts on practicing those skills. Additionally, seeing their scores improve over time can provide learners with

a sense of accomplishment and motivation to continue playing and learning.

2) *Principles of human-computer interaction design:* HCI is a very essential part in an application since it will enable the application to be more successful, safe, useful, and functional. In the long run, it will also make the user's experience more delightful. It is also important to have HCI involved in all stages of product or system development. MMZ also followed the human-computer interaction (HCI) [10][11] which based on Norman's Seven Principles [12][13]. Below are some of the principles that MMZ have applied in the development of the game.

- **Visibility:** Users should be able to tell what their options are and how to access them just by glancing at an interface. This is very important in mobile, web and a game application because making everything visible inside the limited screen space is difficult; therefore, only the options that are required should be included. For example, MMZ provided a very large environment scenery. This is to make sure that the user can see everything in the screen where there is nothing that is hidden from the user. They can also move the mouse around to see more of their surrounding environment.
- **Feedback:** After each action, the user must receive feedback to determine whether their activity was successful. For instance, change the tab's symbol to a spinning wheel to show that a webpage is loading. With the example provided, MMZ applied this in the each of the scene. Every time user entered a scene, the scene name will be listed in the middle top of the screen. This is to make sure that the user will know in what scene they are currently in. Other than that, there is also checkpoint's name listed on top of the checkpoint to help user to know how many checkpoints they have gone through.
- **Affordance:** The link between how things seem and how they're used is called affordability. It is also important for digital apps where the design should be simple enough that users can access the information, they want merely by looking at the interface. For example, MMZ provided an instruction at the very beginning of the game. This is to make sure that user can know what their objectives are before starting to play the game. Other than that, the button in the interface used a familiar design to make sure that user can recognize the button just through the design of the button itself.
- **Constraints:** Constraints is important to prevent a particular form of interaction between an interface and the user. In this example, since the game used the theme of maze, there will be lots of walls where user is unable to go through the wall. User will have to find their way out by finding the correct path to the end of the maze while finding checkpoints and solve the mathematical questions. Other than that, user will also be unable to continue exploring the maze if they did not get the 49-right answer during in the checkpoint scene. They need

to make sure that they get the right answer even if more time is needed to solve the questions. This is to make sure that user can perfectly understand the question and are able to get the right answer.

- Consistency: Learning new patterns is a process that can be used to improve one's abilities. It is also a requirement to avoid frustration when one does not get the same output from the same project. For example, MMZ has the same pattern when the game is going to be played which is that player need to explore the maze while solving the mathematical question provided in the checkpoint. They will return to the maze after solving the questions in the checkpoint scene. Other than that if the player answered the question wrongly, they would have to start all over from the start of the question again. This is to make sure that they are certain and remember the answer for each of the questions.

3) *Instructional design*: The construction of learning experiences and resources in such a way that knowledge and skills are acquired and applied is referred to as instructional design. MMZ also followed instructional design (ID) which based on Mayer's 12 Multimedia Principles [14][15][16].

- The Coherence Principle: Coherence Principle [17] states that humans learn the best when extra and distracting material is not included. For example, MMZ does not have any extra text in the main game scene. User will only have to focus on walking around the maze while answering the question in the checkpoint and finding their way out from the maze.
- The Spatial Contiguity Principle: The Spatial Contiguity Principle [18] refers to the physical distance between the text and visuals on the screen, claiming that relevant information and visuals are best learned when they are physically close together. For example, in the question scene in each of the checkpoint, there will be an image provided with the questions. This is to help the user to answer the questions while referring to the image.
- The Segmenting Principle: Humans learn best when information is delivered in segments rather than in one single continuous stream, according to the Segmenting Principle. Mayer discovered that when students have control over their learning pace, they perform better on memory exams. For example, there are notes provided in the game to help the user to answer the question if they are yet to do any revision. Other than that, students can also repeat themselves answering the questions many times if they have gotten the answer wrongly.
- The Multimedia Principle: According to the Multimedia Principle, humans learn better from words and pictures than from just words. This notion, that visuals plus words are more effective than words alone, is the bedrock of all Mayer's principles. This is the reason why MMZ was developed. It is to help the user to learn mathematics through the game which consists of images, scenery that turns into an educational game.

Instead of learning mathematics in the traditional way, they can make a revision by playing the game where they can repeat many times in answering the questions to help them to become more familiar with the topic.

#### IV. RESULTS

As the pandemic happening, school is unable to be open which lead us to unable to gain an experimental result directly from the student themselves. However, the targeted audience has been changed from primary school students to the public which are differentiated by age. This is because adults can give a better opinion about the game in. Also make sure that the participant is someone who have a very close relation with the student in primary school. This is to ensure that the feedback that the survey will get can be used to improve MMZ. For example, parent, students who have younger siblings in standard 6 etc. There are three age category which is below 20 years old, 20 to 25 years old and 25 years old above. This is to make sure that the results can get different opinion from people so that improvement can be made to MMZ. A survey has been conducted to gain data and analysis for the proposed game.

Through the survey, the survey has managed to gather 19 respondents who are willing to participate and review MMZ. 9 of the participants (47.4%) are at the age of 19 and below, 7 of them (36.8%) are at the age of between 20 and 25 and 3 of them (15.8%) are at the age of 26 and above. The survey is also divided into five sections in which each of them represent different part. Fig. 3 shown below is the age of the respondent.

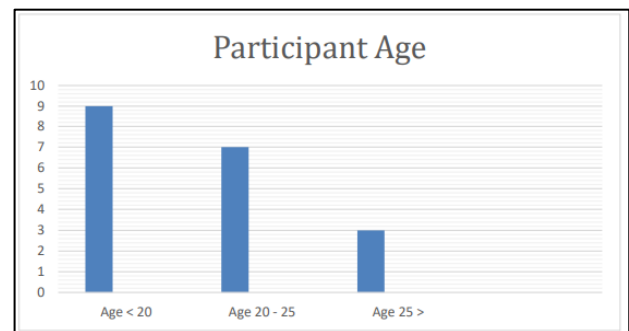


Fig. 3. Age of the respondents

The first section is about the usage of MMZ towards the learning comprehension. From the analysis gathered, 94.7% agreed that MMZ will be able to attract students to learn about mathematics. 89.5% agreed that students will be able to learn mathematics easily through the usage of MMZ. 94.7% agreed that MMZ are a suitable tool for student to do a revision. 94.7% agreed that students can gain more knowledge about mathematics through the usage of MMZ. 89.5% agreed that MMZ will be able to help students to increase their performance in learning mathematics.

The second section is about whether MMZ is a user-friendly application. 94.7% agreed that MMZ are very easy to be played with. 100% agreed that MMZ gameplay are easy to understand. 84.2% agreed that students will be able to learn with ease while playing with MMZ 100% agreed that students can learn while playing with MMZ. 94.7% agreed that students can play MMZ anywhere they want.

The third section is about the procedure of using MMZ. 94.7% agreed that the instruction given in MMZ is detailed and clear. 100% agreed that the language used in MMZ is easy to be understood 100% agreed that MMZ rules are easy to be followed. 100% agreed that MMZ achieved the level of comprehension for a student. 94.7% agreed that students can learn freely by using MMZ without needing a teacher's guidance.

The fourth section is about the usage of MMZ on student's creative and critical thinking. 100% agreed that MMZ can encourage students to think while answering the mathematical question. 100% agreed that MMZ can encourage students to make a discussion among their friends. 89.5% agreed that MMZ can encourage students to design a strategy to achieve success.

The fifth section is about the effectiveness of MMZ Here, the respondent gave an exact percentage of the marks that they can achieve while playing the game. 5.3% gained 40% marks, 36.8% can gain 60% marks, 36.8% are able to achieve 80% marks and 21.1% are able to achieve 100% marks. Overall percentage can see from Table III below.

TABLE III. OVERALL PERCENTAGE FROM THE FEEDBACK TAKEN

Category	Details	Results	
		Yes	No
Learning Comprehension	- Attract students to study	94.7%	5.3%
	- Easily understanding		
	- Reviewing learning	89.5%	10.5%
	- Gain knowledge	94.7%	5.3%
	- Improve learning's performance	94.7%	5.3%
User friendly	- Easy navigating	94.7%	5.3%
	- Easy to understand	100%	0%
	- Making it easier to study	84.2%	15.8%
	- Playing while Learning	100%	0%
	- Study anywhere	94.7%	5.3%
Procedure for MMZ	- Instructions given	94.7%	5.3%
	- Language easy to understand	100%	0%
	- Rules are easy to follow	100%	0%
	- Compatible to learn math	100%	0%
	- Learning independently	94.7%	5.3%
Creative and Critical Thinking	- Encourage to think	100%	0%
	- Encourage social interaction	100%	0%
	- Encourage planning	89.5%	10.5%
Effectiveness	- Learning Math through MMZ is effective		

## V. DISCUSSION

In this section, potential future improvements that have been identified through the analysis will be discussed. The following enhancements can be made to MMZ:

- **Dual-language:** The game should consist of dual language due to the lack of knowledge in term of the languages from the participant. They had issues in understanding the mathematical questions due to not being able to comprehend the meaning of the question.
- **Scoring System:** There should be a scoring system to help the user to understand their level of knowledge instead of them trying to answer the questions multiple times to get the correct answers.
- **More Challenges:** There should be more challenges provided in the game such as number of lives, enemies and traps provided in the game to make the game more challenging and fun to be played instead of them just wandering around the maze.
- **Voice-Over:** There should be a voice-over for each of the question in the checkpoint to help the user to be able to comprehend and understand the question better.
- **Answer Walkthrough:** There should be an answer walkthrough provided for the users who are unable to answer the questions correctly.
- **More Difficulty Level:** There should be more difficulty level according to the age of the students so that more students are able to use MMZ to help them learning mathematics much easier.

## VI. CONCLUSION

MMZ is a comprehensive and interactive educational game that has been specifically designed to assist standard six students in overcoming their difficulties and challenges in learning math subjects. By providing a fun and engaging way to practice and reinforce essential math concepts and skills, this game aims to enhance the students' confidence, motivation, and performance in math, and to support their overall academic success and growth. MMZ also implemented principle of HCI to make sure that the game can improve the ways in which people interact with computers and other digital devices. Instructional designs are also implemented to create effective, efficient, and engaging learning materials and experiences that support the learners' acquisition and retention of knowledge, skills, and attitudes.

The survey results provided valuable feedback about the game, and there are several suggestions for improvement that will make the game more effective as a learning tool for a wider audience. These suggestions include adding a dual language option, a scoring system, more challenges, voice-over, step-by-step explanations of answers, and different difficulty levels. By implementing the recommended suggestions, MMZ has the potential to not only become a more effective educational tool for primary school students, but also for students of other grade levels. While the UPSR exams had been dissolved, standard six students still have a final entrance

exam that will determine their academic standing and help them choose a suitable high school.

#### REFERENCES

- [1] Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897.
- [2] Almanthari, A., Maulina, S., & Bruce, S. (2020). Secondary School Mathematics Teachers' Views on E-Learning Implementation Barriers during the COVID-19 Pandemic: The Case of Indonesia. *Eurasia journal of mathematics, science and technology education*, 16(7).
- [3] Van Dooren, W., De Bock, D., De Wever, B., & Janssens, D. (2019). A Review of Research on the Impact of Conceptual Understanding in Mathematics Education. *Review of Educational Research*, 88(4), 679-715
- [4] Irfan, M., Kusumaningrum, B., Yulia, Y., & Widodo, S. A. (2020). Challenges during the pandemic: use of e-learning in mathematics learning in higher education. *Infinity Journal*, 9(2), 147-158.
- [5] S. Sakulwichitsintu, D. Colbeck, L. Ellis and P. Turner, "A Peer Learning Framework for Enhancing Students' Learning Experiences in Online Environments," 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT), 2018, pp. 168-169, doi: 10.1109/ICALT.2018.00123.
- [6] J Burkholder, EW, Murillo-Gonzalez, G, Wieman, C; Burkholder, Eric W., Murillo-Gonzalez, Gabriel, Wieman, Carl Importance of math prerequisites for performance in introductory physics, 2021, PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH Wieman, Carl/0000-0003-1449-9319; Burkholder, Eric/0000-0001-7420-4290
- [7] A. H. Abdullah, B. Shin, U. H. A. Kohar, D. F. Ali, N. A. Samah and Z. M. Ashari, "A Comparative Study of Teaching Problem-Solving in Mathematics Secondary Schools in Malaysia and South Korea," 2019 IEEE International Conference on Engineering, Technology and Education (TALE), 2019, pp. 1-8, doi: 10.1109/TALE48000.2019.9226011.
- [8] National Council of Teachers of Mathematics (NCTM). (n.d.). Why Is Mathematics Important? ,2021, Retrieved from <https://www.nctm.org/Publications/Why-Is-Mathematics-Important/>
- [9] KPM, (2019), Pelaporan Tafsiran Sekolah Rendah 2019. Kementerian Pelajaran Malaysia
- [10] Al Mahdi, Z., Rao Naidu, V., & Kurian, P. (2019). Analyzing the Role of Human Computer Interaction Principles for E-Learning Solution Design. In *Smart Technologies and Innovation for a Sustainable Future* (pp. 41-44). Springer, Cham.
- [11] AL-Sayid, F., & Kirkil, G. (2022). Students' Web-based activities moderate the effect of human-computer-interaction factors on their e-learning acceptance and success during COVID-19 pandemic. *International Journal of Human-Computer Interaction*, 1-24.
- [12] Hasani, L. M., Sensuse, D. I., & Suryono, R. R. (2020, September). User-centered design of e-learning user interfaces: A survey of the practices. In *2020 3rd International Conference on Computer and Informatics Engineering (IC2IE)* (pp. 1-7). IEEE.
- [13] McCarthy, S., Rowan, W., Ertiö, T., Lynch, L., & Kahma, N. (2021). Open e-learning platforms and the design-reality gap: exploring the impact of user-perceived functional affordances.
- [14] Mayer, R. E. (2019). How multimedia can improve learning and instruction.
- [15] Hamdani, M. (2023). Analysis of e-content of Khuzestan Province teachers during the COVID-19 period based on Mayer's principles. *International Journal of Technology Enhanced Learning*, 15(1), 95-103.
- [16] Oh, E. G., Chang, Y., & Park, S. W. (2020). Design review of MOOCs: Application of e-learning design principles. *Journal of Computing in Higher Education*, 32(3), 455-475.
- [17] Werdiningsih, T., Triyono, M. B., & Majid, N. W. A. (2019). Interactive multimedia learning based on mobile learning for computer assembling subject using the principle of multimedia learning (Mayer). *International Journal of Advanced Science and Technology*, 28(16), 711-719.
- [18] Mahajan, R., Gupta, K., Gupta, P., Kukreja, S., & Singh, T. (2020). Multimedia instructional design principles: Moving from theoretical rationale to practical applications. *Indian Pediatrics*, 57(6), 555-560.

# An Improved Poisson Surface Reconstruction Algorithm based on the Boundary Constraints

Zhouqi Liu<sup>1</sup>, Lei Wang<sup>\*2</sup>, Muhammad Tahir<sup>3</sup>, Jin Huang<sup>4</sup>, Tianqi Cheng<sup>5</sup>, Xinping Guo<sup>6</sup>, Yuwei Wang<sup>7</sup>, ChunXiang Liu<sup>\*8</sup>

School of Computer Science and Technology, Shandong University of Technology, Zibo, 255000, China<sup>1,2,4,5,6,7</sup>  
Department of Computer Science, Mohammad Ali Jinnah University, Block 6 PECHS, Karachi, 75400, Pakistan<sup>3</sup>  
School of Resources and Environmental Engineering, Shandong University of Technology, Zibo, 255000, China<sup>8</sup>

**Abstract**—The usage of the point cloud surface reconstruction to generate high-precision 3D models has been widely applied in various fields. In order to deal with the problems of insufficient accuracy, pseudo-surfaces and high time cost caused by the traditional surface reconstruction algorithms of the point cloud data, this paper proposes an improved Poisson surface reconstruction algorithm based on the boundary constraints. For large density point cloud data obtained from 3D laser scanning, the proposed method firstly uses an octree instead of the KD-tree to search the near neighborhood; then, it uses the Open Multi-Processing (OpenMP) to accelerate the normal estimation based on the moving least squares algorithm; meanwhile, the least-cost spanning tree is employed to adjust the consistency of the normal direction; and finally a screened Poisson algorithm with the Neumann boundary constraints is proposed to reconstruct the point cloud. Compared with the traditional methods, the experiments on three open datasets demonstrated that the proposed method can effectively reduce the generation of pseudo-surfaces. The reconstruction time of the proposed algorithm is about 16% shorter than that of the traditional Poisson reconstruction algorithm, and produce better reconstruction results in the term of quantitative analysis and visual comparison.

**Keywords**—Point cloud; moving least squares; Poisson reconstruction; Neumann boundary constraints

## I. INTRODUCTION

With the popularity of 3D reconstruction technology, the surface reconstruction has been widely applied in various fields, such as the mapping [1], driverless [2], medical technology [3], robotics [4]. For the methods of reconstructing surfaces based on the 3D point cloud data, they can be generally divided into two schemes: the local methods and the global methods. The local reconstruction methods divide all the point clouds into small blocks of data, which are then reconstructed locally and finally connect them together by using some kinds of stitching functions [5, 6]. This scheme retains the surface texture features well, but it is heavily sensitive to the noise. The global surface reconstruction is an approximation of all the point clouds, and the implicit surface is optimally reconstructed by solving extreme values and other methods [7]. It is under a high overall smoothness and is

suitable for the interpolation and the whole repair of irregular, non-uniform scattered data. As an implicit function-based surface reconstruction method, Poisson reconstruction method reconstructs a model with watertight closure features, geometric surface properties and detail characteristics [8]. After normal estimation on the input point cloud data, the objective of this algorithm is to estimate the indicator function of the model and extract the isosurface, and then complete the surface reconstruction using the moving cube algorithm based on the indicator function and isosurface [9]. Compared to the other algorithms, it allows a hierarchy of local basis functions to be divided and the reconstruction is projected to be a Poisson space problem, combining both of the advantages of the global and the local methods.

Recently, a large number of good works on the Poisson reconstruction has been reported by the famous scholars. For example, Kazhdan et al. proposed a screened Poisson surface reconstruction algorithm to keep the sparse point structure, and solved it by using a multiple mesh algorithm [10]. Z. Xu et al. proposed an adaptive bandwidth Gaussian kernel density estimator that facilitates the removal of the noise and anomalies in the 3D reconstruction process [11]. F. Gao et al. proposed a Poisson reconstruction algorithm based on the improved isosurface extraction, which can effectively eliminate the problems of surface holes and disconnect the surface features [12]. B. Ummerhofer et al. proposed a generalized convolutional kernel for 3D reconstruction with ConvNets from point clouds [13]. Though the above reconstruction methods have made great breakthroughs in this field, there still are some drawbacks to overcome. For example, the traditional Poisson reconstruction is prone to pseudo surface and normal inconsistencies. The time complexity of reconstruction method for improved isosurface extraction is too high. The method based on an adaptive bandwidth Gaussian kernel density estimator is not effective for the reconstruction of point cloud data with small density variation. Considering the entire above problem to deal with, this paper proposes an improved Poisson reconstruction algorithm and reduces the reconstruction time of the model to improve the reconstruction efficiency.

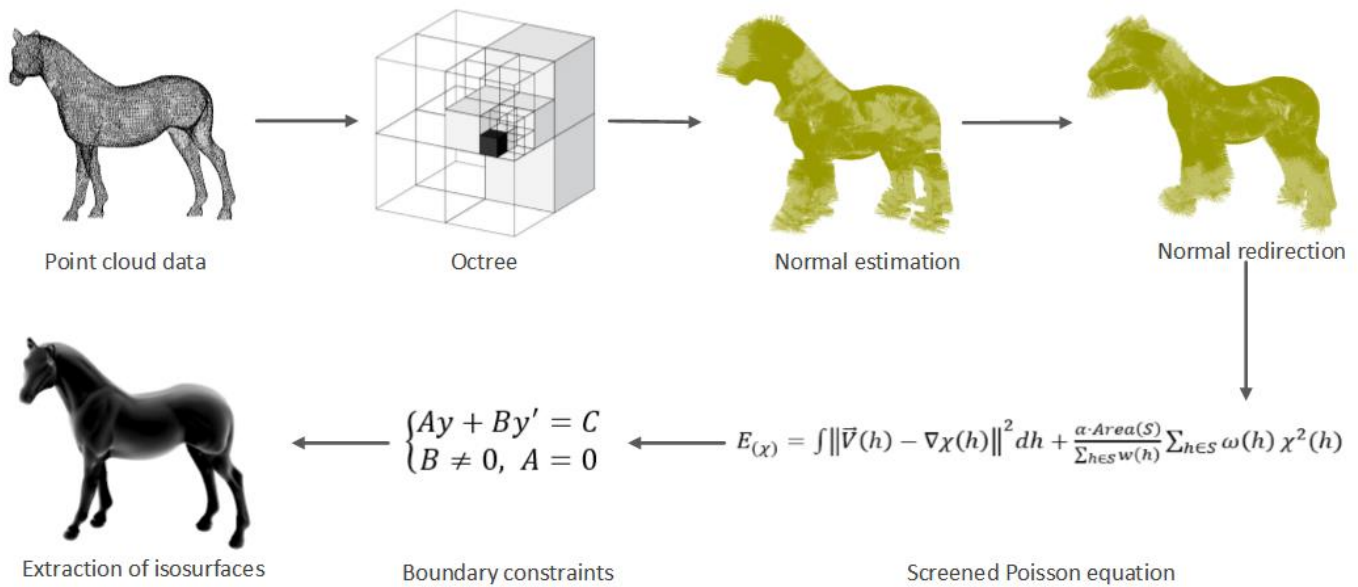


Fig. 1. The flowchart of the improved reconstruction algorithm

The remainder of this paper is organized as follows. Section II presents the proposed method of reconstruction for 3D point cloud. In Section III, the experiments are implemented to evaluate the method. Conclusions are given in Section IV.

## II. WHOLE METHOD

In Fig. 1, the proposed improved reconstruction algorithm consists the following steps: the set of gradients of the indicator function is determined by the normal vector of the point cloud data, so the accuracy of the normal estimation has a significant impact on the surface reconstruction results. Compared to the traditional Poisson algorithm which mostly uses principal component analysis for normal estimation, this paper proposes an improved method. Firstly, an octree is used instead of a KD-tree to search the nearest neighborhood; then the normal of the point cloud is estimated by moving least squares and accelerated by OpenMP [14], and then the normal direction is adjusted consistently by a least-cost spanning tree. The traditional Poisson reconstruction algorithm is prone to generate pseudo-surfaces. In this paper, the screened Poisson algorithm is implemented by introducing constraints on the location and gradient of the points to constrain the surface reconstruction process; and adding Neumann boundary constraints to make the solution of the indicator function more accurate and generate more accurate surface models. The whole flow of the specific algorithm is shown in Fig. 1.

### A. Octree and the Moving Least Squares

The octree is used to represent a three-dimensional space, and is a spatial extension of the quadtree [15]. The geometric entities in the three-dimensional space are first dissected into cubes, each with the same time and space complexity, and then the geometric objects in the three-dimensional space of size  $(2n*2n*2n)$  are dissected by a circular recursive partitioning method to construct a directional graph with a root node [16]. In the octree structure if the cubes being divided have the same properties, the cube forms a leaf node [17]; otherwise the

dissection of the cube into eight sub-cubes continues[18]. For  $(2n*2n*2n)$  size space objects are dissected at most  $n$  times, and the structure is shown in Fig. 2 [19]. The octree algorithm is more efficient than the KD-tree in searching point cloud data with high data volumes and is more automated, allowing optimization of the time for 3D reconstruction of point clouds [20, 21].

The Moving Least Squares algorithm is simple and easy to implement as a method for interpolating discrete data, with high fitting accuracy [22]. When a large amount of discrete data is distributed in a heterogeneous manner, the use of traditional least squares algorithms often requires the fitting of segments to the data, in addition to avoiding the problem of discontinuous and unsmooth fitting curves on adjacent segments [23]. Moving least squares introduces the concept of tight branches and the fitting function is constructed differently.

1) *The fitting function:* The fitting function [24] for the moving least squares in the local area is shown in Eq. (1).

$$(x) = \sum_{i=1}^n \beta_i(x)q_i(x) = q^T(x)\beta(x) \quad (1)$$

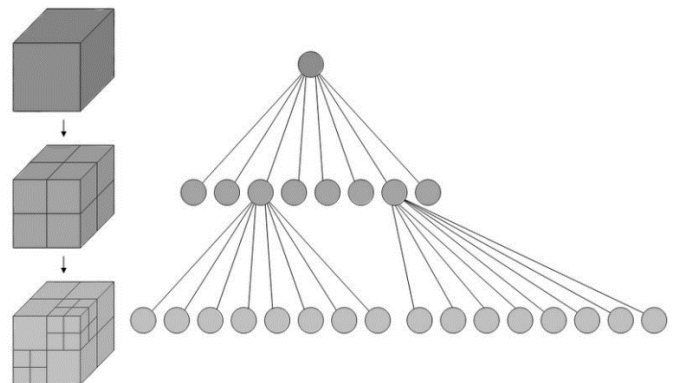


Fig. 2. The flowchart of the improved reconstruction algorithm

where  $\beta(x) = [\beta_1(x), \beta_2(x), \dots, \beta_n(x)]^T$  is the coefficient to be found;  $q(x) = [q_1(x), q_2(x), \dots, q_n(x)]^T$  is a polynomial of order  $k$  as a basis function; and the number of terms in the basis function is  $n$ . The basis functions for the linear and quadratic bases are  $q(x) = [1, x, y]^T$  and  $q(x) = [1, x, y, x^2, xy, y^2]^T$ .

In this paper, the quadratic basis is used to improve the accuracy of surface fitting and normal estimation [25]. The squared weight of the difference between the local approximation of  $f(x_i)$  and the nodal value  $y_i$  is minimized, the weighted residual equation is shown in Eq. (2).

$$J = \sum_{i=1}^m w(x - x_i) * [F(x) - y_i]^2 = \sum_{i=1}^m w(x - x_i) * [q^T(x_i)\beta(x) - y_i]^2 \quad (2)$$

where  $m$  is the number of nodes in the area of influence;  $w(x - x_i)$  is the weight function of node  $x_i$ . In order to determine the coefficient  $\beta(x)$ , it must be made to take a minimal value. The derived equation is shown in Eq. (3).

$$\frac{\partial J}{\partial \beta} = A(x)\beta(x) - B(x)y = 0 \quad (3)$$

where  $A(x) = \sum_{i=1}^m w(x - x_i)q(x_i)q^T(x_i)$  ;  $B(x) = [w(x - x_1)q(x_1), w(x - x_2)q(x_2), \dots, w(x - x_m)q(x_m)]$  ;  $y^T = [y_1, y_2, \dots, y_m]$ .

The coefficients of the fitting equation are shown in the following equation.

$$\beta(x) = A^{-1}(x)B(x)y \quad (4)$$

Substitute Equation (4) into Eq. (1) and let  $\varphi^k(x) = [\varphi_1^k, \varphi_2^k, \dots, \varphi_m^k] = q^T(x)A^{-1}(x)B(x)$ , the fitted function of the moving least squares is obtained as shown in Eq. (5).

$$f(x) = \sum_{i=1}^m \varphi_i^k(x)y_i = \varphi^k(x)y \quad (5)$$

where  $\varphi^k(x)$  is the shape function;  $k$  is the order of the basis function.

2) *The selection of the weight function:* The weight function  $w(x - x_i)$  in Eq. (2) is compactly supported in the moving least squares algorithm, that is, the weight function is only affected by the subdomain near  $x$ , which is called the influence region of point  $x$ . Beyond this region, its weight is small and the influence can be ignored. As an indispensable part of the moving least square method, the selection of weight function is very important to the fitting accuracy [26,27]. Generally speaking, the circle is chosen as the support domain of the weight function, and its radius is denoted as  $S_{max}$ , as shown in Fig. 3.

The common weight functions include the spline weight function and the Gaussian weight function [28,29]. When Gaussian weight function is used for the moving least squares surface fitting, the width of a single kernel is difficult to meet the feature requirements of the whole model. It needs to be selected according to the point cloud density and curvature distribution in the local neighborhood, and the kernel width is not easy to be determined. As reported in literature [30], the cubic spline weight function has continuity. Therefore, the cubic spline weight function is selected in this paper. Denoting

$s = x - x_i$ ,  $\bar{s} = s/s_{max}$ , the cubic spline weight function is shown as follows:

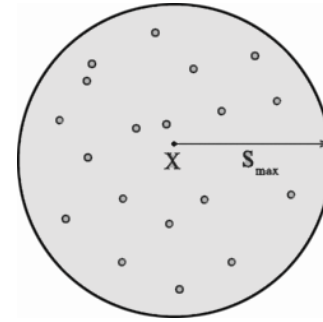


Fig. 3. The area of the impact region

$$w(\bar{s}) \begin{cases} \frac{2}{3} - 4\bar{s}^2 + 4\bar{s}^3 & |\bar{s}| \leq \frac{1}{2} \\ \frac{4}{3} - 4\bar{s} + 4\bar{s}^2 - \frac{4}{3}\bar{s}^3 & \frac{1}{2} < |\bar{s}| \leq 1 \\ 0 & |\bar{s}| > 1 \end{cases} \quad (6)$$

### B. The Normal Estimation and Direction Consistency

Define a local quadratic surface as  $S(r, v) = S(r, v, h(r, v))$ . The first order partial derivative of the surface at a point is then calculated as shown in the following equation.

$$\begin{cases} S_r = (1, 0, 2ar + bv + e) \\ S_v = (0, 1, br + 2cv + f) \end{cases} \quad (7)$$

where  $h(r, v) = ar^2 + brv + cv^2 + dr + ev + f$ . So the normal direction of the surface at this point is calculated by the following formula. Where,  $a, b, c, d, e$  and  $f$  are the coefficients of the surface equation.

$$n = \frac{S_r * S_v}{|S_r * S_v|} = \frac{(-2ar - bv - e, -br - 2cv - f, 1)}{\sqrt{(2ar + bv + e)^2 + (br + 2cv + f)^2 + 1}} \quad (8)$$

The normal direction calculated based on the moving least square method is ambiguous, that is, only the line where the normal line is located is obtained, but the direction of the normal line is not determined. Therefore, it is necessary to carry out direction consistency processing for the normal in the model to ensure its accuracy [31]. In this paper, the minimum spanning tree (MST) is used to adjust the normal direction uniformly [32,33]. Firstly, the algorithm defines a cost function for the point cloud in the model, as shown in the following equation.

$$cost(p * q) = |g * n_p + g * n_q| \quad (9)$$

Where  $g$  is the unit vector pointing from point  $p$  to point  $q$ .  $p$  and  $q$  are adjacent points, and  $n_p$  and  $n_q$  are normal to points  $p$  and  $q$ , respectively.

### C. Screening Factors

Because Poisson equation is a partial differential equation, some errors will cause the indicator function to shift, resulting in a large number of pseudo-closed surfaces at the edges of non-watertight surface connections [34,35]. In this paper, the process of surface reconstruction is constrained by introducing the position and gradient constraints, and then the screened Poisson reconstruction algorithm is implemented.



In the process of solving the Poisson equation, the indicator function  $\chi: R^3 \rightarrow R$  is solved for the directed point set  $\vec{V}: R^3 \rightarrow R^3$ , so that the gradient of the function and the directed point set form the best approximation, that is, the scale function is minimized to solve the problem  $\min_{\chi} \|\nabla\chi - \vec{V}\|$ ,  $E_{(\chi)} = \int \|\nabla\chi(h) - \vec{V}(h)\|^2 dh$ . In this paper, a point set  $S: S \rightarrow R \geq 0$  with weight  $\omega$  is given, and some gradient constraints and value constraints of discrete points are added to minimize the scale function to recalculate the sample point function. The constraint equation is shown in Eq. (10).

$$E_{(\chi)} = \int \|\vec{V}(h) - \nabla\chi(h)\|^2 dh + \frac{\alpha \cdot Area(S)}{\sum_{h \in S} \omega(h)} \sum_{h \in S} \omega(h) \chi^2(h) \quad (10)$$

where,  $h$  is the sample point of the input point set,  $\nabla\chi$  is the gradient of the indicator function,  $Area(S)$  is the surface region to be reconstructed,  $\alpha$  is the screening factor, which is used to measure the fitting gradient and the fitting value, and  $\omega(h)$  is the sample point weight. For the convenience of calculation, the weight  $\omega(h)$  of each sample point is set as 1 in this paper. Eq. (11) can be obtained by simplified to be:

$$E_{(\chi)} = \langle \vec{V} - \nabla\chi, \vec{V} - \nabla\chi \rangle_{[0,1]^3} + \alpha(\chi, \chi)_{(\omega, S)} \quad (11)$$

where,  $\langle \cdot, \cdot \rangle_{[0,1]^3}$  represents the standard inner product on the space of functions in the unit cube (scalar and vector values).  $\alpha(\chi, \chi)_{(\omega, S)}$  is the representation of the unit cube on the function space. This value is obtained from the weighted summation of the functions of the sampling points, as shown in Equation (12).

$$\langle F, G \rangle_{(\omega, S)} = \frac{\alpha \cdot Area(S)}{\sum_{h \in S} \omega(h)} \sum_{h \in S} \omega(h) F(h)G(h) \quad (12)$$

#### D. The Neumann Boundary Constraint

In order to satisfy the solution of the shielded Poisson equation and the boundary conditions in a given region, the algorithm bias error is reduced by introducing boundary constraints for solving the partial differential equation. There are three main types of boundary constraints: the Dirichlet boundary constraint [36], the Neumann boundary constraint and the Robin boundary constraint [37,38]. At the endpoints it is generally written in the form  $Ay + By' = C$ .

a) If  $B = 0$  and  $A \neq 0$ , it is called the Dirichlet boundary condition, also called the first type of boundary condition, and gives the corresponding value of the unknown function on the boundary.

b) If  $B \neq 0$ ,  $A = 0$ , then it is called Neumann boundary condition, also called the second type of boundary condition, and gives the directional derivative of the unknown function normal to the boundary outside.

c) If  $B \neq 0$ ,  $A \neq 0$ , then it is called Robin boundary condition, also called the third type of boundary condition, and gives a linear combination of the value of the function of the unknown function on the boundary and the directional derivative of the outer normal.

The Neumann boundary constraint is added to the Poisson reconstruction process to force the normal derivative of the implicit function to be zero at the boundary, which makes the

solution of the indicator function more accurate. Both the Dirichlet and Neumann boundary constraints require the implicit function to take a negative value outside the model. Since the gradient of the indicator function is equal to the vector field, Neumann boundary conditions are more conducive to the solution of the indicator function than Dirichlet. Neumann boundary conditions constrain the watertight surface and require the derivative of the implicit function to be zero on the boundary of the integral domain. This property is compatible with the gradient of the indicator function, since the vector field  $V$  is numerically set to zero in the region far from the point cloud sample. When the surface exceeds the boundary of the integral domain, the Neumann boundary constraint creates a deviation across the boundary of the domain.

### III. EXPERIMENTAL RESULTS AND DISCUSSION

To verify the effectiveness, accuracy and feasibility of the proposed algorithm, it is compared with the traditional Poisson reconstruction algorithm and greedy projection triangulation algorithm [39,40]. The 3D scanning datasets from Stanford University and GeometryHub are used in this paper. All experiments are performed on a computer with a quad-core AMD Ryzen 3 3100 CPU and 16GB RAM configured at 3.59 GHz. And the parallel acceleration is implemented by the OpenMP.

#### A. Results

In Fig. 4, the original point cloud data of Model 1, Model 2 and Model 3 are shown from left to right. Fig. 5 shows the normal estimation of the MLS algorithm for the three data respectively, and the calculated normal is ambiguous. Fig. 6 shows the initial normal is estimated by the moving least squares algorithm for three data, and then the minimum spanning tree is used to redirect the normal. Based on the point cloud data in Fig. 6, the proposed algorithm is compared with the traditional Poisson reconstruction algorithm, the greedy projection triangulation algorithm.

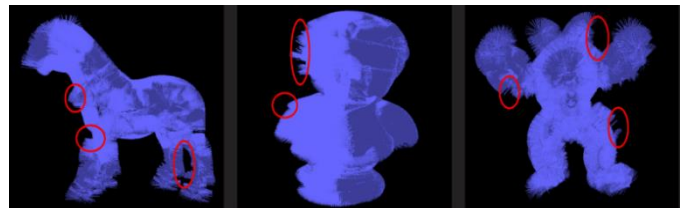


Fig. 4. The normal before redirection



Fig. 5. The normal after redirection

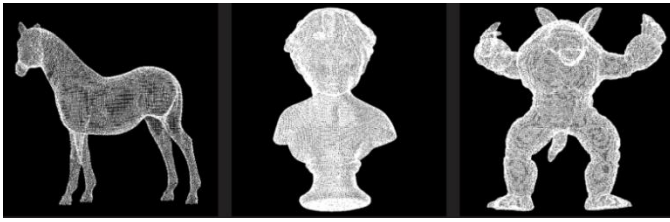


Fig. 6. The original point cloud data

In Fig. 7, Fig. 7(a), 7(d) and 7(g) show the model generated by the traditional Poisson reconstruction algorithm, Fig. 7(b), 7(e) and 7(h) show the model generated by the greedy projection triangulation algorithm, and Fig. 7(c), 7(f) and 7(i) show the model generated by the proposed algorithm in this paper.

In order to further optimize the reconstruction effect, the different parameters of the proposed algorithm are analyzed and compared with the point cloud data of the above models. The effect diagram of model 1 is shown in Fig. 8. Then the above point cloud data were reconstructed with three types of boundary conditions respectively. The experimental results are shown in Fig. 9 and Table II. Fig. 9(a), 9(d) and 9(g) is the surface reconstruction under Neumann boundary constraints, while Fig. 9(b), 9(e) and 9(h) is the surface reconstruction under Dirichlet boundary constraints. Fig. 9(c), 9(f) and 9(i) shows the surface reconstruction under Robin boundary constraints.

### B. Discussions

It can be seen from Fig. 5, the normal direction is confused, someone pointing to the inner side, and someone pointing to the outer side. If there is no redirection, further processing will lead to many reconstruction errors. Fig. 6 shows the use of normal redirection optimizes the consistency of the normal direction and provides a more accurate normal input for later reconstruction. As can be seen from the comparison of several groups of models in Fig. 7, the surface features of the traditional Poisson algorithm have poor sealing property, and it is easy to generate pseudo surfaces by misconnecting the regions that do not belong to morphological features at the edges, among which Fig. 7(g) is the most obvious. The model reconstructed by greedy projection triangulation algorithm has many small holes, and the model surface is rough. Compared with the former, the proposed algorithm can effectively solve the whole problem by introducing shielding factor and boundary constraint. And the overall model is more perfect in reduction and surface smoothness.

Based on the data in Table I, the reconstruction time of the traditional Poisson reconstruction algorithm, greedy projection triangulation algorithm and the proposed algorithm is compared. The reconstruction time of the proposed algorithm is about 16% shorter than that of the traditional Poisson reconstruction algorithm, but the reconstruction time is longer than that of the greedy projection triangulation algorithm. This is because the proposed algorithm uses the moving cube algorithm to extract the isosurface, which consumes a long time and needs further improvement. When calculating the point cloud model with larger data volume, the algorithm in this paper has higher reconstruction efficiency.

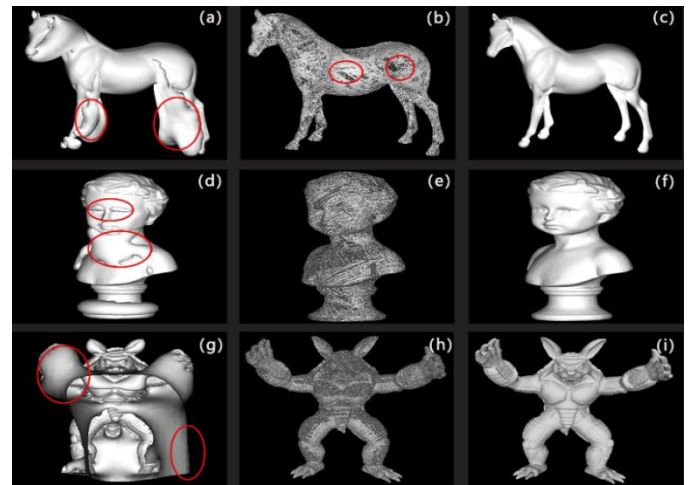


Fig. 7. The surface reconstruction results based on different algorithms

In Fig. 8, the higher the value of the screening factor, the more detailed the reconstructed model will be, but the longer it will take. At the screening factor of 4, the detailed features are well represented and take a moderate amount of time. As can be seen from Fig. 9, the accuracy of the reconstruction model under Dirichlet constraint is low, and many detailed features are lost. Although Robin boundary constraint is a combination of Dirichlet boundary constraint and Neumann boundary constraint, it takes the longest time and has a poor reconstruction effect on the edge, as shown in the Fig. 9(c) for the pseudo-surface marked by the red circle. The reconstruction time under the constraint of Neumann boundary is the shortest. Compared with the traditional Poisson surface reconstruction, the pseudo-surface is significantly reduced, and the effect is more ideal.

TABLE I. THE COMPARISON OF RECONSTRUCTION EFFECTS BASED ON DIFFERENT ALGORITHMS

The experimental data	Traditional Poisson reconstruction algorithm			Greedy projection triangulation algorithm			The proposed algorithm		
	Vertices	Patches	Time/s	Vertices	Patches	Time/s	Vertices	Patches	Time/s
Model 1	99723	199446	12.6	51424	102848	5.1	82571	165138	9.8
Model 2	379188	758372	50.2	300643	601374	23.8	356274	712543	42.1
Model 3	548685	1097370	76.6	446261	892541	35.4	523165	1046297	63.3

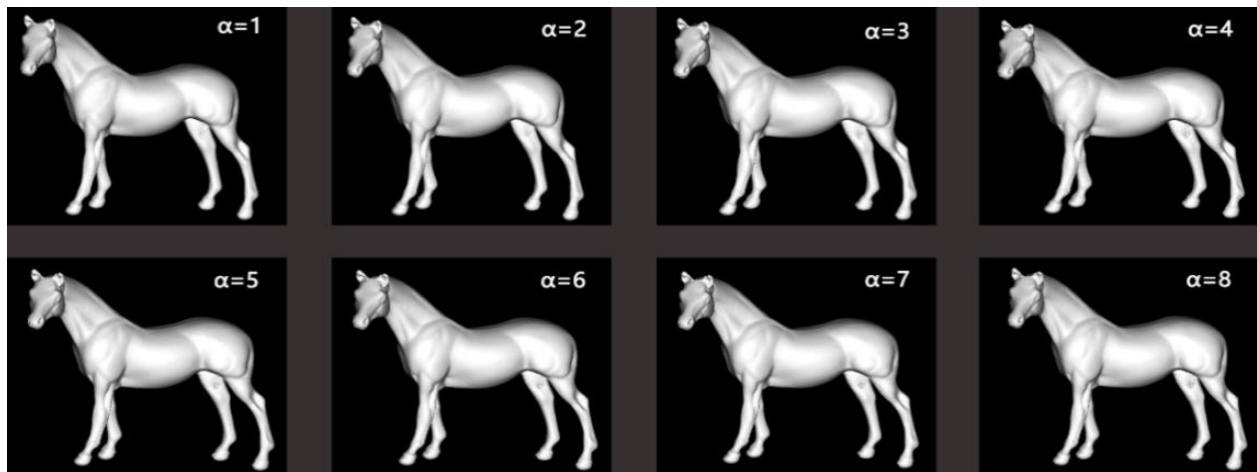


Fig. 8. The surface reconstruction with different screening factors

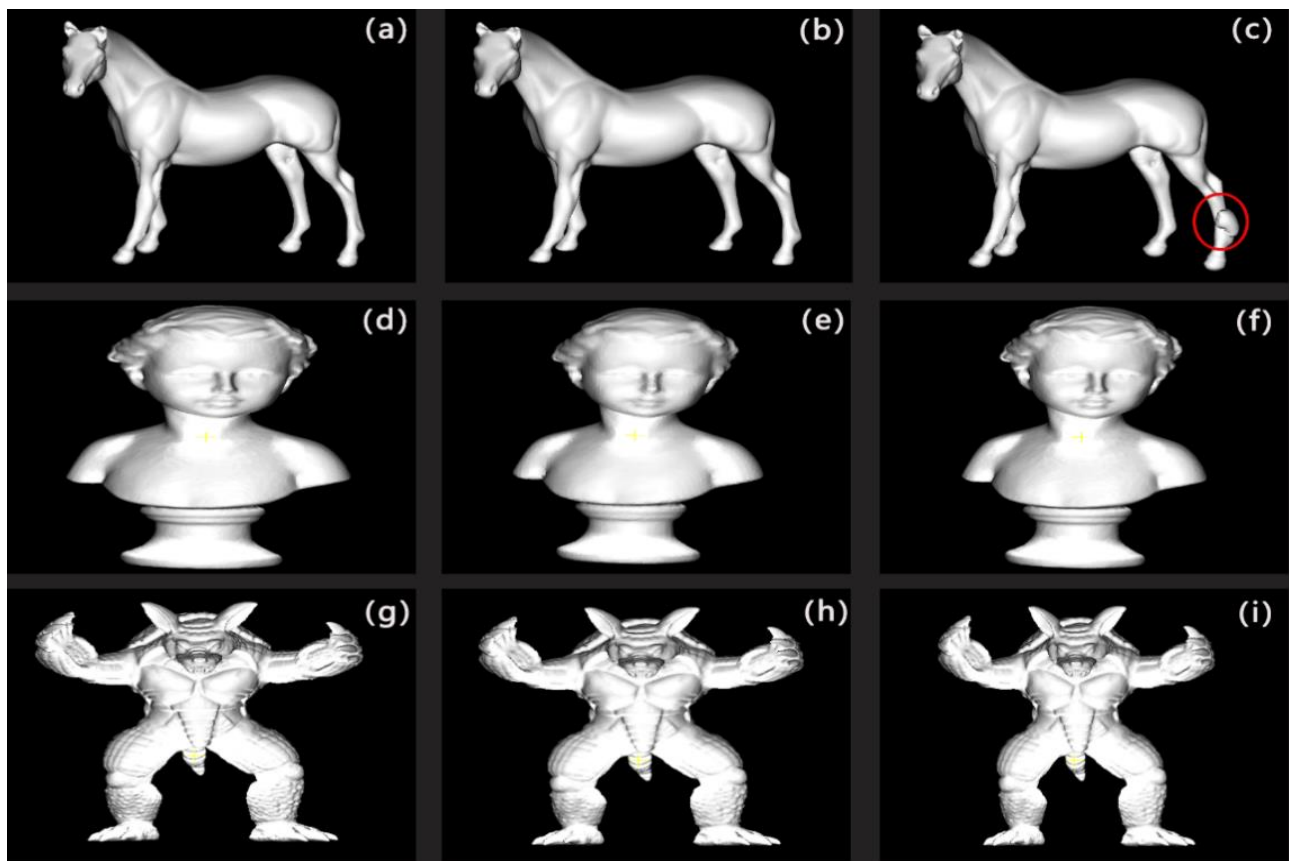


Fig. 9. The Surface reconstruction with different boundary constraints

TABLE II. THE COMPARISON OF RECONSTRUCTION EFFECTS BASED ON DIFFERENT BOUNDARY CONSTRAINTS

The experimental data	Neumann			Dirichlet			Robin		
	Vertices	Patches	Time/s	Vertices	Patches	Time/s	Vertices	Patches	Time/s
Model 1	82571	165138	9.8	83626	167248	10.2	84509	169082	10.9
Model 2	356274	712543	42.1	358823	717646	45.4	359716	719432	46.8
Model 3	523165	1046297	63.3	529371	1058741	66.9	530805	1061610	68.3

#### IV. CONCLUSION

In this paper, we propose an improved Poisson surface reconstruction algorithm based on the Boundary Constraints. Firstly, octree is used to replace KD-tree for nearest neighbor search. Secondly, the normal vector is estimated by moving least square method, and the redirection based on minimum cost spanning tree is used to reduce the error. Finally, on the basis of traditional Poisson reconstruction, screening factor and Neumann boundary constraint are introduced to further improve the reconstruction effect.

The experimental results on different data show that the proposed algorithm achieves more accurate reconstruction results, which can effectively reduce the generation of pseudo-surfaces and also reduce the running time to a certain extent. The further work is to improve the extraction of isosurface on the basis of the proposed algorithm, and try to apply the modified algorithm to other fields to obtain high-quality reconstructed models.

#### ACKNOWLEDGMENT

This study was supported by A project ZR2021MF017 supported by Shandong Provincial Natural Science Foundation; A project ZR2020MF147 supported by Shandong Provincial Natural Science Foundation; A project 2020SNPT0055 supported by SDUT&Zibo City Integration Development Project; The National Natural Science Foundation of China (61502282).

#### REFERENCES

- [1] M. Puliti, G. Montaggioli, A. Sabato, "Automated subsurface defects' detection using point cloud reconstruction from infrared images," *Automation in Construction*, Vol. 129, pp. 103829, 2021
- [2] D. Y. Duan, W. G. Qiu, Y. J. Cheng, Y. C. Zheng, F. Lu, "Reconstruction of shield tunnel lining using point cloud," *Automation in Construction*, Vol. 130, pp. 103860, 2021
- [3] Q. Cheng, P. Sun, C. Yang, Y. Yang, P. X. Liu, "A morphing-Based 3D point cloud reconstruction framework for medical image processing," *Computer methods and programs in biomedicine*, Vol. 193, pp. 105495, 2020
- [4] J. Kim, D. Chung, Y. Kim, H. Kim, "Deep learning-based 3D reconstruction of scaffolds using a robot dog," *Automation in Construction*, Vol. 134, pp. 104092, 2022
- [5] J. Choe, S. Im, F. Rameau, M. Kang, I. S. Kweon, "Volumefusion: Deep depth fusion for 3d scene reconstruction," *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 16086-16095, 2021
- [6] Q. Xu, W. Wang, D. Ceylan, R. Mech, U. Neumann, "Disn: Deep implicit surface network for high-quality single-view 3d reconstruction," *Advances in Neural Information Processing Systems*, Vol. 32, 2019
- [7] S. Cuomo, A. Galletti, G. Giunta, L. Marcellino, "Reconstruction of implicit curves and surfaces via RBF interpolation," *Applied Numerical Mathematics*, Vol. 116, pp. 157-171, 2017
- [8] S. L. Liu, H. X. Guo, H. Pan, P. S. Wang, X. Tong, Y. Liu, "Deep implicit moving least-squares functions for 3D reconstruction," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 1788-1797, 2021
- [9] S. Peng, C. Jiang, Y. Liao, M. Niemeyer, M. Pollefeys, A. Geiger, "Shape as points: A differentiable poisson solver," *Advances in Neural Information Processing Systems*, Vol. 34, pp. 13032-13044, 2021
- [10] M. Kazhdan, H. Hoppe. "An Adaptive Multi - Grid Solver for Applications in Computer Graphics," *Computer graphics forum*. Vol. 38, No. 1, pp. 138-150, 2019
- [11] Z. Xu, C. Xu, J. Hu, Z. Meng, "Robust resistance to noise and outliers: Screened Poisson Surface Reconstruction using adaptive kernel density estimation," *Computers & Graphics*, Vol. 97, pp. 19-27, 2021
- [12] F. Gao, H. Zhou, C. Huang. "Improved Poisson Reconstruction Algorithm Based on Vector Field and Isosurface," *Laser&Optoelectronics Progress*, Vol. 57, No. 10, pp. 101016, 2020
- [13] B. Ummenhofer, V. Koltun, "Adaptive surface reconstruction with multiscale convolutional kernels," *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 5651-5660, 2021
- [14] S. Bak, C. Bertoni, S. Boehm, R. Budiardja, B. M. Chapman, J. Doerfert, P. K. Yeung, "OpenMP application experiences: Porting to accelerated nodes," *Parallel Computing*, Vol. 109, pp. 102856, 2022
- [15] Y. T. Su, J. Bethel, S. Hu, "Octree-based segmentation for terrestrial LiDAR point cloud data in industrial applications," *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 113, pp. 59-74, 2016
- [16] S. Han, "Towards efficient implementation of an octree for a large 3D point cloud," *Sensors*, Vol. 18, No. 12, pp. 4398, 2018
- [17] M. Gouda, J. Mirza, J. Wei, A. Ribeiro Castro, K. ElBasyouny, "Octree-based point cloud simulation to assess the readiness of highway infrastructure for autonomous vehicles," *Computer-Aided Civil and Infrastructure Engineering*, Vol. 36, No. 7, pp. 922-940, 2021
- [18] M. Bassier, M. Bonduel, B. Van Genechten, M. Vergauwen, "Segmentation of large unstructured point clouds using octree-based region growing and conditional random fields," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 42, pp. 25-30, 2017
- [19] Y. Chen, L. Zhou, Y. Tang, J. P. Singh, N. Bouguila, C. Wang, J. Du, "Fast neighbor search by using revised kd tree," *Information Sciences*, Vol. 472, pp. 145-162, 2019
- [20] R. Pinkham, S. Zeng, Z. Zhang, "Quicknn: Memory and performance optimization of kd tree based nearest neighbor search for 3d point clouds," *2020 IEEE International Symposium on High Performance Computer Architecture (HPCA)*. IEEE, pp. 180-192, 2020
- [21] W. Hou, D. Li, C. Xu, H. Zhang, T. Li, "An advanced k nearest neighbor classification algorithm based on KD-tree," *2018 IEEE International Conference of Safety Produce Informatization (IICSPI)*. IEEE, pp. 902-905 2018
- [22] W. Y. Tey, N. A. Che Sidik, Y. Asako, M. Muhiudeen, O. Afshar, "Moving least squares method and its improvement: A concise review," *Journal of Applied and Computational Mechanics*, Vol. 7, No. 2, pp. 883-889, 2021
- [23] V. Mohammadi, M. Dehghan, "A divergence-free generalized moving least squares approximation with its application," *Applied Numerical Mathematics*, Vol. 162, pp. 374-404, 2021
- [24] J. G. Manchuk, C. V. Deutsch, "Boundary modeling with moving least squares," *Computers & Geosciences*, Vol. 126, pp. 96-106, 2019
- [25] B. Sober, D. Levin, "Manifold approximation by moving least-squares projection (MMLS)," *Constructive Approximation*, Vol. 52, No. 3, pp. 433-478, 2020.
- [26] L. Zhang, F. Guo, H. Zheng, "The MLS-based numerical manifold method for nonlinear transient heat conduction problems in functionally graded materials," *International Communications in Heat and Mass Transfer*, Vol. 139, pp. 106428, 2022.
- [27] W. Q. Wang, Y. Yan, G. R. Liu, "An IB-LBM implementation for fluid-solid interactions with an MLS approximation for implicit coupling," *Applied Mathematical Modelling*, Vol. 62, pp. 638-653, 2018
- [28] A. Mohtashami, A. Akbarpour, M. Mollazadeh, "Development of two-dimensional groundwater flow simulation model using meshless method based on MLS approximation function in unconfined aquifer in transient state," *Journal of Hydroinformatics*, Vol. 19, No. 5, pp. 640-652, 2017
- [29] Y. Hilali, O. Bourihane, "A mixed MLS and Hermite-type MLS method for buckling and postbuckling analysis of thin plates," *Structures*. Elsevier, Vol. 33, pp. 2349-2360, 2021
- [30] E. Bertolazzi, M. Frego, F. Biral, "Point data reconstruction and smoothing using cubic splines and clusterization," *Mathematics and Computers in Simulation*, Vol. 176, pp. 36-56, 2020.

- [31] D. Lu, X. Lu, Y. Sun, J. Wang, "Deep feature-preserving normal estimation for point cloud filtering," *Computer-Aided Design*, Vol. 125, pp. 102860, 2020.
- [32] J. Yang, M. Hao, X. Liu, Z. Wan, N. Zhong, H. Peng, "FMST: an automatic neuron tracing method based on fast marching and minimum spanning tree," *Neuroinformatics*, Vol. 17, No. 2, pp. 185-196, 2019
- [33] S. Du, R. Lindenbergh, H. Ledoux, J Stoter, L Nan, "AdTree: accurate, detailed, and automatic modelling of laser-scanned trees," *Remote Sensing*, Vol. 11, No. 18, pp. 2074, 2019
- [34] A. Maiti, D. Chakravarty, "Performance analysis of different surface reconstruction algorithms for 3D reconstruction of outdoor objects from their digital images," *SpringerPlus*, Vol. 5, No. 1, pp. 1-26, 2016
- [35] Q. Xin, Z. He, "Three dimensional stratum interpolation and visualization based on section and borehole data from jointing the moving least square method and poisson reconstruction method," *Earth Science Informatics*, Vol. 13, No. 4, pp. 1341-1349, 2020
- [36] M. Kazhdan, M. Chuang, S. Rusinkiewicz, H. Hoppe, "Poisson surface reconstruction with envelope constraints," *Computer graphics forum*, Vol. 39, No. 5, pp. 173-182, 2020
- [37] A. Sefer, A. Yapar, "Reconstruction algorithm for impenetrable rough surface profile under Neumann boundary condition," *Journal of Electromagnetic Waves and Applications*, Vol. 36, No. 8, pp. 1154-1172, 2022
- [38] D. Bochkov, F. Gibou, "Solving Poisson-type equations with Robin boundary conditions on piecewise smooth interfaces," *Journal of Computational Physics*, Vol. 376, pp. 1156-1198, 2019
- [39] S. K. Gupta, D. P. Shukla, "Application of drone for landslide mapping, dimension estimation and its 3D reconstruction," *Journal of the Indian Society of Remote Sensing*, Vol. 46, No. 6, pp. 903-914, 2018
- [40] Y. Zhang, H. Wu, W. Yang, "Forests growth monitoring based on tree canopy 3D reconstruction using UAV aerial photogrammetry," *Forests*, Vol. 10, No. 12, pp. 1052, 2019

# Data Augmentation for Deep Learning Algorithms that Perform Driver Drowsiness Detection

Ghulam Masudh Mohamed, Sulaiman Saleem Patel\* and Nalindren Naicker  
Department of Information Systems, Durban University of Technology  
Durban, South Africa

**Abstract**— Driver drowsiness is one of the main causes of driver-related motor vehicle collisions, as this impairs a person's concentration whilst driving. With the enhancements of computer vision and deep learning (DL), driver drowsiness detection systems have been developed previously, in an attempt to improve road safety. These systems experienced performance degradation under real-world testing due to factors such as driver movement and poor lighting. This study proposed to improve the training of DL models for driver drowsiness detection by applying data augmentation (DA) techniques that model these real-world scenarios. This paper studies six DL models for driver drowsiness detection: four configurations of a Convolutional Neural Network (CNN), two custom configurations as well as the architectures designed by the Visual Geometry Group (VGG) (i.e. VGG16 and VGG19); a Generative Adversarial Network (GAN) and a Multi-Layer Perceptron (MLP). These DL models were trained using two datasets of eye images, where the state of eye (open or closed) is used in determining driver drowsiness. The performance of the DL models was measured with respect to accuracy, F1-Score, precision, negative class precision, recall and specificity. When comparing the performance of DL models trained on datasets with and without DA in aggregation, it was found that all metrics were improved. After removing outliers from the results, it was found that the average improvement in both accuracy and F1 score due to DA was +4.3%. Furthermore, it is shown that the extent to which the DA techniques improve DL model performance is correlated with the inherent model performance. For DL models with accuracy and F1-Score  $\leq 90\%$ , results show that the DA techniques studied should improve performance by at least +5%.

**Keywords**—Data augmentation; deep learning; computer vision; drowsiness detection; road safety

## I. INTRODUCTION

Road accidents represent a major socio-economic challenge for individuals, industries, and nations [1]. Commuters involved in road accidents are affected in a variety of ways; such as death, sustaining physical injuries, psychological trauma, as well as incurring financial burdens from damage to property [1-4]. For industries, road accidents adversely affect supply chain performance and logistics, reducing operational efficiency [5-7]. The net result of this adversely impacts the economy of a country. Furthermore, for national authorities, road accidents cause traffic congestion; resulting damage to infrastructure and increased environmental pollution. Road accidents are a greater concern in developing countries, wherein more than 90% of accidents result in fatalities [1]. Of all developing countries, the World Health Organisation

reports that South Africa has the poorest road safety record, with approximately 14 000 deaths per annum and an accident fatality rate of 3.2% [2, 8, 9].

The factors that cause road accidents need to be identified before an effective solution can be developed. Studies, such as those presented by Machetele and Yessoufou [1] and Verster and Fourie [2], highlight that driver-related accidents account for 80% to 90% of fatal road accidents. A key cause of driver-related accidents is drowsiness (which may result from excessive alcohol consumption), as this impairs a person's concentration and focus [2, 10]. The detection of driver fatigue or drowsiness is hence essential towards improving road safety and reducing the accident rate [11, 12].

In light of the fourth industrial revolution, technology is becoming more ubiquitous and there is growing motivation to utilize artificial intelligence and machine learning to solve social problems, such as driver drowsiness detection. To this end, there have been a range of studies that apply deep learning (DL) techniques to solve the problem of driver drowsiness detection [13-19]. DL is a subset of machine learning that mimics the neural network of the human brain, thus creating an artificial neural network [14]. Artificial neural networks comprise of multiple nodes that model neurons of the human brain, which are organized into layers [20]. Data is propagated from the input layer to the output layer. These artificial neural networks have the potential to solve regression and classification problems, including image classification problems [20, 21]. In the context of image classification, each layer trains upon the output of the previous layer, enabling latter layers to identify more intricate elements of the images [21].

At a technical level, the aforementioned studies perform driver drowsiness detection by considering images of a driver's eye, and using DL algorithms to determine the eye state (i.e. whether the eye is opened or closed). By applying this technology to frames from a video feed of the driver, it is possible to determine whether eyes are closed for extended periods of time, which is an indicator of drowsiness. Some of the DL algorithms used in literature include: (i) convolutional neural networks (CNNs) of different configurations [14-16, 18, 22, 23]; (ii) the multi-layer perceptron (MLP) [13, 24]; (iii) the respective Visual Geometry Group 16 (VGG16) [25, 26] and 19 (VGG19) [17, 26] models; as well as (iv) the generative adversarial network (GAN)[27]. The reported accuracies of the models in these studies range between 75% and 96%.

Despite the high accuracies reported in the studies, real-world challenges during implementation were reported that adversely affected the accuracy of the trained models. Among these challenges were: (i) poor lighting, where lighting is either too bright or too dim [13, 14, 17, 19]; (ii) changes to the driver's seat position [22]; (iii) a change in the angle of the driver's face while driving [13, 22] the use of spectacles and/or sunglasses by drivers [14, 17-19, 24].

In this paper, the authors proposed to address these real-world challenges by performing data augmentation (DA) on the training image sets that are input into DL models for driver drowsiness detection. DA techniques introduce artificial images that simulate real-world effects [28], such as different lighting environments and changes to face orientation. This study also uses a training dataset containing images of drivers with and without eyewear to address the challenges associated with drivers wearing spectacles or sunglasses. The DA techniques are tested on CNN models, GAN models, MLP models and both the VGG16 and VGG19 models. Hyperparameter tuning is performed on all models to optimize their learning rate and enhance their overall performance. Literature has shown that careful selection of hyperparameters has a significant impact on model performance [28, 29]. The effect of the DA is evaluated by comparing the performance of models trained with and without DA in with respect to the following metrics: (i) accuracy, (ii) precision, (iii) negative class precision, (iv) recall, (v) specificity, and (vi) F1-score. It is hypothesized that the use of DA will result in improved performance of all models.

It is noted that previous studies in literature [14, 25, 27] have incorporated the use of DA in improving the performance of their specific driver drowsiness detection models. However, to the best of the authors' knowledge, there are no comprehensive studies that investigate DA techniques for a wide range of DL algorithms in the context of driver drowsiness detection, as is done in this paper.

The research in this paper makes the following contributions:

- 1) Presenting an overview of DA techniques to model the specific real-world scenarios that cause challenges for driver drowsiness detection systems.
- 2) Studying the DA techniques on a wide range of DL models that perform driver drowsiness detection and statistically analyzing the effects of the DA techniques.
- 3) Demonstrating the extent to which the DA techniques studied are able to improve DL models that perform driver drowsiness detection and proposing a design guideline for DL model developers on that conditions under which the DA techniques should be considered.

The rest of this paper is organized as follows. In Section II, a review of existing literature was presented. Section III presents the materials and methods used in this study, including providing an overview of a real-world drowsiness detection system. In Section IV the results of the investigations are presented and finally, conclusions and insights that were drawn from this study are presented in Section V. Section V also makes recommendations for future work.

## II. RELATED WORK

This section reviews the DL algorithms that have been extensively used in previous studies, to implement models and applications, for drowsiness detection in motorists.

A study by Jabbar et al. [14] proposed a drowsiness detection system that could be implemented on the driver's dashboard, using an Android phone. The system was able to predict the drowsiness of the driver based on their eye state. This study made use of a CNN network to implement a binary classification model that was able to classify the drowsiness in facial images. Data augmentation techniques were applied to the images, before they were trained on the model. The Dlib C++ library was used to extract the driver's facial landmarks from the images. These facial features were fed into the algorithm for training. The dataset was created using the extracted eye features. This model achieved an accuracy of 83.3%. A similar study by Zhang, Su, Geng and Xiao [18] was conducted to detect the drowsiness of a person, using the eye state. This proposed model was implemented on an Infrared video camera. The AdaBoost algorithm was used to extract facial landmarks from the images. The extracted eye landmarks were used to create the image dataset, to train the model on. The CNN model was used as the binary classifier for drowsiness. An accuracy of 95.8% was achieved by this study.

Sharan, Viji, Pradeep and Sajith [15] proposed a similar drowsiness detection system to Jabbar et al. [14] that could be implemented on the driver dashboard. However, this study proposed that a Raspberry Pi camera module be used to capture the drivers face. The drowsiness prediction was also based on the eye state. The Haar Cascade classifier was used for facial extraction during the implementation of this system and the CNN network was implemented as the binary classifier. Contrast Level Adaptive Histogram Equalization was applied to remove the noise and improve the picture quality, before they were trained on the CNN model. The CNN model was trained on an existing dataset, comprising of eye images. The study by Seetharman, Sridhar and Mootha [22] made use of a CNN network to classify the drowsiness in images. The prediction was based on the eye and mouth state of the extracted faces. The Dlib library was utilized to extract the facial regions from the images, similar to the study done by Jabbar et al. [14]. A dataset for the model was then generated using the extracted eye regions. The trained CNN model achieved an accuracy of 92.4%. In addition, this proposed model was intended to be implemented on a dashboard video camera. Chirra, Uyyala and Kolli [16] proposed a similar model for drowsiness detection, as a CNN network was used to predict the drowsiness in images. The eye state was the metric for prediction, with the Viola-Jones algorithm used to extract the facial landmarks from the images, during the implementation of this system. An existing dataset of eye images was used to train the CNN model. The model produced an accuracy of 96.42%. This model was also proposed to be implemented on a video camera for drowsiness detection, like the study conducted by Seetharman, Sridhar and Mootha [22].

A model using the VGG 19 model to detect driver drowsiness, based on the eye state, was proposed by Hashemi, Mirrashid and Shirazi [17]. This study made use of the Viola-

Jones algorithm to extract the facial landmarks from the images. The extracted eye landmarks were then used to create the dataset for this model. The Viola-Jones algorithm has been utilised in previous work [16]. This model obtained an accuracy of 94.96%, with its intended application in driver dashboard monitoring. A study by Ahuja, Saurav, Srivastava and Shekhar [26], proposed an approach to improved drowsiness detection, by using a knowledge distillation technique to reduce the size of DL models, whilst maintaining high accuracy. A large model will have high memory consumption and longer response times. Therefore, there was a need to reduce the size of the DL model. The Histogram of Gradient algorithm was used to extract the facial regions from the images, during system implementation. VGG19 and Visual Geometry Group 16 (VGG16) were the algorithms used to train their respective models, to classify the drowsiness in images. These models were trained on an existing dataset, consisting of eye images. The predictions were based on the eye state for both models. The VGG19 and VGG16 models, obtained the accuracy of 92.5% and 95% respectively.

Bajaj, Ray, Shedge, Jaikar and More [25] proposed a real-time drowsiness prediction system that will be implemented on an Android application, to monitor the driver's face from the dashboard. This system can predict the drowsiness using the driver's eye state. A comparative analysis of three DL algorithms, specifically: Inception, ResNet-50 and VGG 16 were performed. Data augmentation techniques were applied to the images, before they were trained on the models. The models were trained on an existing dataset, comprising of face images. The accuracy achieved by the Inception, ResNet-50 and VGG 16 models were 89%, 56% and 91%, respectively.

A study by Jabbar et al. [13] proposed a system for drowsiness detection that could be implemented on an android application, for dashboard monitoring. The prediction of this system was based on the driver's eye state. The Dlib C++ library was used to extract the person's facial landmarks from the images. This library has been used for facial feature extraction in previous work [14,25]. These facial features were used to create the dataset, which was fed into the MLP algorithm for training. The model was able to classify a driver as either drowsy or non-drowsy. An accuracy rate of 80.92% was achieved by this model. A similar study by Ghourabi, Ghazouani and Barhoumi [24] made use of the MLP algorithm to detect drowsiness in the images. The eye and mouth state were used to classify the drowsiness. The Histogram of Gradient algorithm was used to extract the facial regions from the images. These extracted facial regions were used to create the dataset that was fed into the model for training. The model is intended to be implemented for dashboard monitoring. This study obtained an accuracy rate of 74.9%.

Ngxande, Tapamo and Burke [27] proposed a framework to reduce the biasness of a model during the training process. A Generative Adversarial Network (GAN) model was trained on an image dataset. This model made predictions using facial landmarks and the eye state in particular. The extracted facial landmarks were used to create the dataset for model training. Data augmentation techniques were applied to the images before they were loaded into the GAN model. This helped to improve the performance of the binary classification model. An accuracy rate of 91.62% was achieved by this model.

Many of the studies have used facial and eye extraction algorithms, to create image datasets from real-time data, to train their models on. However, this study aimed to use existing datasets that were available online, to train the DL models. The reason for this was because, this study aimed on improving the performance of trained models, regardless of the source of data. Therefore, no facial and eye extraction algorithms were used on real-time data, in this study.

Literature has shown that many drowsiness detection models faced issues with prediction accuracy, due to poor lighting and the use of sunglasses [13,14,17,18,24]. The other challenge that affected accuracy was the positioning of the driver's face [13,22]. Another gap identified is the lack of pre-processing and data augmentation applied on the data before training. Data augmentation was used in [14,25,27], to create more comprehensive models that exhibits improved performance. DA was used to remove biasness from the models, thus improving the performance. However, not many of the previous studies have comprehensively studied DA to model real-world scenarios to improve model performance, on a wide range of DL algorithms that detect driver drowsiness, as done in this study.

Therefore, this study aimed to develop an improved approach towards drowsiness detection by using data augmentation. Data augmentation techniques were used to create training data that replicate real-life scenarios that correlate with the challenges faced in previous studies.

### III. MATERIALS AND METHODS

This section first provides an overview of a real-world driver drowsiness detection system and isolates the role of the DL algorithms that this study focuses on. The data sources and DA techniques utilized in this paper are then discussed. Thereafter, a technical summary of the DL algorithms considered is provided, along with the parameters used in this study. Finally, the authors present the different evaluation metrics that are used to quantify the performance of the DL algorithms.



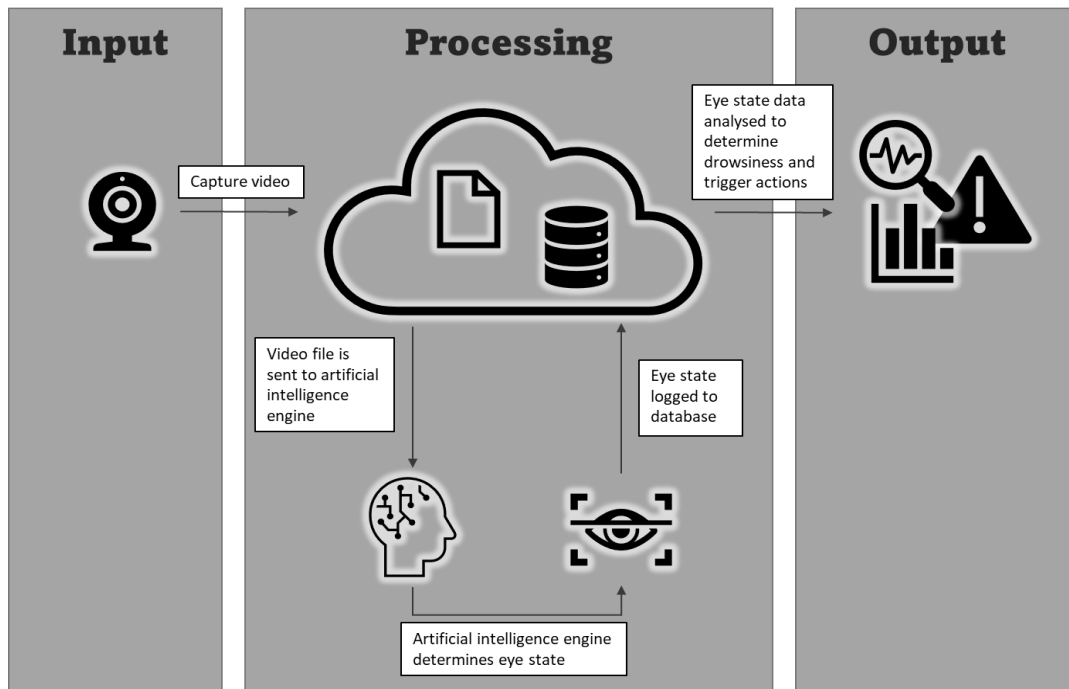


Fig. 1. Overview of a real-world driver drowsiness detection system

#### A. An Overview of a Real-World Drowsiness Detection System

Fig. 1 illustrates the process flow for a real-world driver drowsiness detection system. The process starts with a camera that captures a video of the driver's face, which serves as the input to the system. The camera can either be mounted to the dashboard or steering wheel of the vehicle. The captured video is then stored on cloud-hosted infrastructure, typically in some form of unstructured blob storage.

At the start of the processing stage of the system, the video file is passed on to an artificial intelligence engine, consisting of three sub-units. The first sub-unit extracts individual frames from the video file, which will then be treated as a series of sequential images. The second sub-unit uses image detection techniques to isolate the eye from each image of the driver's face. This produces a series of sequential images of the driver's eyes. Finally, the third sub-unit utilizes a pre-trained DL model to analyze the images and determine the state of the driver's eye (open or closed) in each frame. The eye state determined in each frame is then logged in a database, which is also typically cloud-hosted.

In the final stage of the system, the eye states stored in the database are analyzed and interpreted to detect the drowsiness of the driver. Drowsiness detected when the driver's eyes are in the 'closed' state for extended periods (multiple consecutive frames from the video feed).

#### B. Design and Configuration of Study

The research presented in this study focuses on the third sub-unit of the artificial intelligence engine, viz. the DL

algorithm that determines the driver's eye state, as described in Section III.A. Hence, for the experiments conducted, the inputs in this study were images of a driver's eye and the outputs were a categorical variable indicating the eye state. A binary categorical output was used, with the positive class label indicating the "open" eye state and the negative class label indicating the "closed" eye state. The experimental configuration used is depicted in Fig. 2.

In performing the experiments, appropriate datasets of eye images were first sourced. In selecting the datasets, the authors ensured that images where the eye was partially obscured by eyewear (spectacles or sunglasses) were included. By doing this, the DL models would learn to distinguish between eye states irrespective of the use of eyewear.

The datasets were then split into training and testing data using an 80:20 ratio. A copy of the training dataset was created, and data augmentation techniques were performed to model the real-world challenges of eye orientation and lighting conditions. Two DL models were trained: one was trained on the original (pre-treatment) training dataset, and the other was trained on the modified (post-treatment) training dataset. Depending on the architecture of the DL algorithm being investigated, any necessary data-shaping modifications were made to the images from the dataset.

The pre-treatment and post-treatment DL models were applied to the testing dataset to evaluate and compare their performance. As was the case with the training datasets, any modifications to the testing dataset required by the DL model architecture were made.

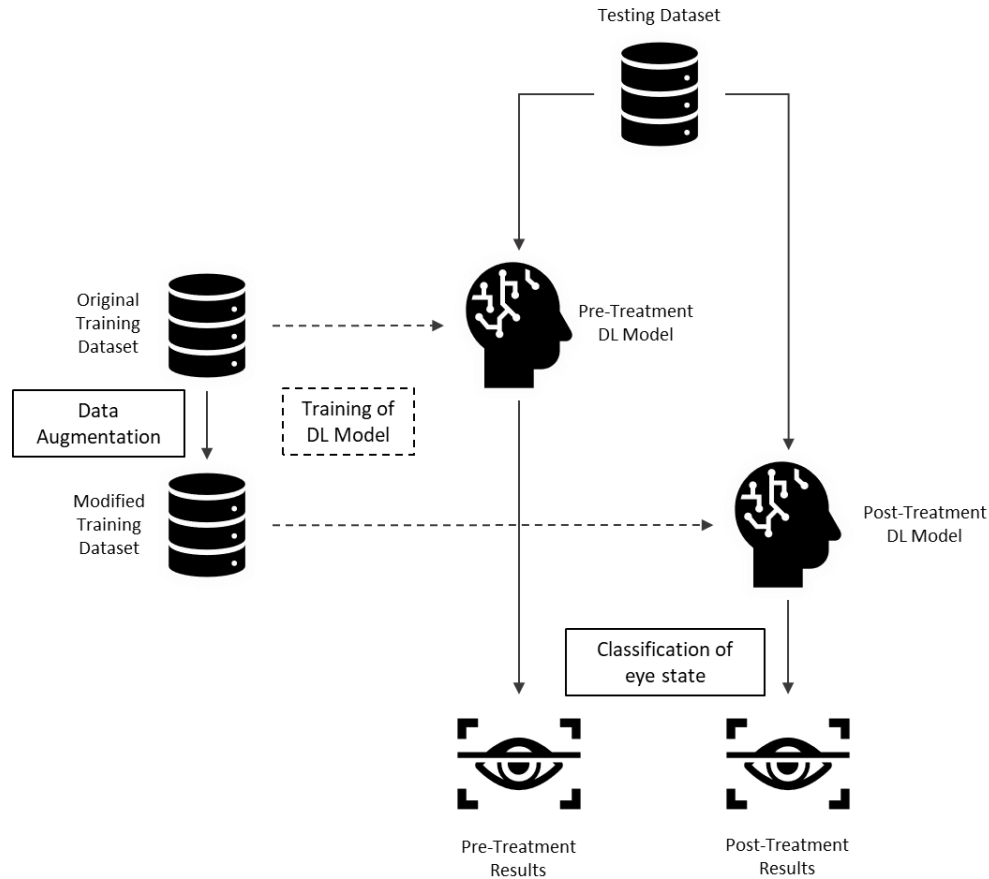


Fig. 2. Configuration of study

TABLE I. PROPERTIES OF DATASETS

Property	Dataset 1	Dataset 2
<b>Total number of images</b>	2 500	1452
<b>Eye Opened</b>	1 250	726
<b>Eye Closed</b>	1 250	726
<b>Image size (pixels)</b>	96 × 96	256 × 256
<b>Colour/Greyscale</b>	Greyscale	Colour
<b>File format</b>	Portable Network Graphics (PNG)	Joint Photographic Experts Group (JPEG)
<b>File compression</b>	Lossless compression	Lossy compression

The experiments were done using pre-built Python libraries on the Jupyter Notebooks development environment. A personal computer equipped with 8 gigabytes of random-access memory, an Intel Core i5-7200U processor and a 64-bit Windows 10 operating system.

1) *Selection of datasets*: There were two datasets utilised in this study, which were obtained from online repositories [30, 31]. Both datasets contained images of human eyes with and without eyewear, and images labelled according to the eye state. The properties of the datasets are presented in Table I.

The balanced distribution of eye states was preserved when splitting each of the datasets into respective training and testing

datasets, using an 80:20 ratio. The Scikit-learn Python library was used to implement the data splitting.

When exploring the datasets, it was also noted that both sets of data contained images from a diverse range of ethnicities. Different skin tones and complexions were noted, as well as different eye shapes. The authors further observed that among female eyes, the extent to which make-up such as eyeliner and false eyelashes were used differed.

2) *Data augmentation and pre-processing*: Data augmentation improves model performance by generating variations of training data [14]. This reduces overfitting and improves the model’s ability to make generalizations [14, 32]. The specific augmentations performed in this study were designed to simulate real-world scenarios and overcome some of the challenges indicated in literature.

The ImageDataGenerator class within the Keras library for Python [33] was used to implement pre-processing and DA in this study. The ImageDataGenerator class supports DA in real-time and makes sure that the model is trained with different variations of images during each training iteration (epoch) [34, 35].

The following pre-processing and data augmentation techniques were applied:

a) *Brightness adjustment*: Multiple studies in literature have shown that poor lighting conditions had a negative impact on the accuracy of DL models for driver drowsiness detection [13, 14, 17, 19]. While driving, ambient lighting conditions can change due to environmental conditions such as the time of day and the weather. For example, driving at night results in a very low brightness conditions and driving in bright sunshine results in very high brightness conditions. While driving, it is also possible for lighting conditions to change rapidly, such as when driving under a bridge/overpass on a sunny day or through the shadow cast by a building or other large structure.

To model scenarios with different lighting environments, this study applied a randomized change to image brightness when augmenting images. This is implemented through adding a constant,  $\delta$ , to all pixels in the image. The brightness adjustment function is mathematically described as:

$$f_b(p) = \max(0, \min(255, p + \delta)) \quad (1)$$

In (1),  $p$  is the value of the pixel and falls in the range  $0 \leq p \leq 255$ . Positive values of  $\delta$  increase the brightness of the image while negative values of  $\delta$  decrease the brightness of the image. The  $\max(\cdot)$  and  $\min(\cdot)$  functions ensure that the brightness-adjusted pixel value remains in a valid range.

b) *Horizontal flips*: The shape of a human eye may differ slightly between the left eye and the right eye. Creating artificial data by flipping the horizontal orientation allows the DL model to be trained to analyze either eye of the driver.

c) *Rotation, translation and zoom*: Literature showed that changes to the driver's face orientation was a real-world scenario that adversely affected the performance of DL models [13, 21]. Therefore, in this study, rotation, translation shifts and zoom transformations were used to model changes to the driver's face orientation. Rotation and translational shifts are useful to simulate movement of a driver's head while travelling. Zoom transformations model a change in depth between the camera and the driver's face, which may result from the driver changing their seat position or posture.

d) *Normalization, centering and standardization*: Normalization and standardization improve the learning rate and reduces the number of epochs required to train a DL model [36, 37]. These processes ensure that no individual input pixel dominates performance [38]. This is done by mathematically adjusting data such that it follows a Gaussian distribution with zero mean and unit variance [39].

Normalization involves rescaling the value of pixels to have a unit maximum, which reduces the computational power required to train the DL model. As all pixels have the same maximum value ( $p = 255$ ), the normalization function is described by [36]:

$$f_n(p) = \frac{p}{255} \quad (2)$$

Centering ensures that the data has a mean of zero, while standardization ensures that the data has a unit variance [36]. Setting these statistical properties of the data improves the rate at which a DL algorithm converges when training, as well as increasing model accuracy by eliminating statistical bias.

Centering and standardization can be applied to data in with respect to individual images (sample-wise) or with respect to the entire set of images (feature-wise). The functions for sample-wise centering (sc), feature-wise centering (fc), sample-wise standardization (ss) and feature-wise standardization (fs) are [39]:

$$f_{sc}(p) = p - \bar{p}_I \quad (3)$$

$$f_{fc}(p) = p - \bar{p}_D \quad (4)$$

$$f_{ss}(p) = \frac{p}{\sigma_I} \quad (5)$$

$$f_{fs}(p) = \frac{p}{\sigma_D} \quad (6)$$

In (2) – (6),  $\bar{p}$  represents the mean pixel value and  $\sigma$  represents the standard deviation of pixel values. The subscripts 'I' and 'D' respectively denote statistics calculated over pixels from a single image (I) and statistics calculated over the entire dataset (D).

In this study, each of the above pre-processing operations is performed on input data.

3) *Deep learning algorithms*: As discussed in Section I, DL is a subset of machine learning and involves mimicking the human brain. DL algorithms follow a common structure, to the extent that they adopt a layered architecture with multiple nodes at each layer. The DL algorithms for this study are designed to perform a binary classification in determining whether the eye state is 'opened' or 'closed'. A brief overview of the different DL algorithms implemented in this study for image classification is provided below.

a) *Convolutional neural network (CNN)*: The CNN is the most popular artificial neural network (at the time of writing). There are typically three classes of layers in a CNN: convolution layers, pooling layers and fully-connected layers [16, 40]. Fig. 3, re-produced from [41], illustrates the layout of these layers.

Convolution and pooling layers work together to perform feature extraction from the input image [16, 40]. First, input data representing pixels of an image is multiplying the kernel filters of a convolution layer to generate feature maps. Thereafter, a pooling layer is used to group features together and reduce the size of the feature maps. Pooling features together improves the computation time of the DL algorithm [16].

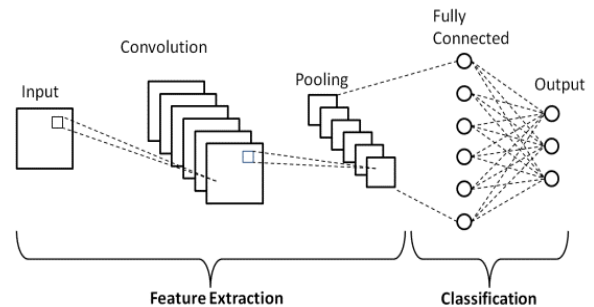


Fig. 3. Basic CNN architecture [41]

TABLE II. ARCHITECTURAL LAYERS OF CNN-C1

Layer Number	Properties
1	Convolutional Layer. 32 nodes, same padding, 3×3 kernel
2	Max pooling layer
3	Convolutional Layer. 32 nodes, same padding, 3×3 kernel
4	Max pooling layer
5	Convolutional Layer. 64 nodes, same padding, 3×3 kernel
6	Max pooling layer
7	Flatten layer
8	Dense layer with 128 units
9	Dense Softmax output layer with two units

TABLE III. ARCHITECTURAL LAYERS OF CNN-C2

Layer Number	Properties
1	Convolutional Layer. 8 nodes, same padding, 3×3 kernel
2	Average pooling layer
3	Convolutional Layer. 16 nodes, same padding, 3×3 kernel
4	Average pooling layer
5	Flatten layer
6	Dense layer with 120 units
7	Dense layer with 84 units
8	Dense Softmax output layer with two units

The processed feature maps are then fed into one or more fully-connected layers. The final layer is referred to as the output layer, and any fully-connected layers between the pooling layer and the output layer are referred to as hidden layers. Each node in a fully-connected layer performs a mathematical operation on its input data using an activation function. These activation functions are selected to map inputs to suitable outputs and perform classification [42].

Two different CNN model configurations were investigated in this study. For brevity, they are referred to as CNN-C1 and CNN-C2. Their respective architectures are shown in Table II and Table III.

Table II describes the first CNN architecture used in this study. These layers are arranged sequentially in a linear stack [43]. The first two convolution layers in this model have 32 nodes each, which are responsible for learning multiple spatial patterns and features from the input image [44]. The last convolution layer 64 nodes. A 3×3 kernel filter is used in each convolution layer, to generate the feature maps. Each convolution layer applied same padding to the input image, which enabled the image to get completely covered by the kernel filter, to generate a feature map [45]. Furthermore, each convolution layer was followed by a pooling layer that applies a maximum filter (max pooling). Once the convolution was completed, the data was then passed to the flatten layer to flatten the multi-dimensional feature map into one dimension

[46]. This single dimensional array was then forwarded into the dense layer of the network. A dense layer of 128 units is then used to perform the image classification, using the output from the convolution layers [47]. The last layer of this network was a two-unit output layer which made use of a softmax activation function that calculated the probabilities of each class [48]. There are only two units used in the output layer, because these models are binary classifiers, with predictions made for only two class labels. The output produced by the softmax layer, is represented in the form of a vector, which contains the probabilities of each class, for every sample data

In addition, a Rectified Linear Unit (ReLU) activation function was added to each convolution layer and dense layer, to ensure no negative values were passed to the subsequent layers [16]. The ReLU activation function is given by:

$$f_{\text{ReLU}}(x) = \max(0, x) \quad (7)$$

In (7),  $x$  refers to the input data to the activation function.

Table III describes the second CNN configuration used in this study, which also consists of sequential layers. This configuration uses fewer convolution layers than CNN-C1, but more fully-connected layers when performing classification. CNN-C2 also applies an averaging filter in the pooling layers (average pooling). As with CNN-C1, a ReLU activation function was added to each convolutional layer and dense layer, to ensure no negative values propagated through the network.

*b) Visual geometry group (VGG) networks 16 and 19:* The VGG have conducted extensive research into DL algorithms for image classifications that improve upon the traditional CNN [49]. The two VGG algorithms chosen were VGG16 [50] and VGG19 [51]. The VGG16 model consists of 13 convolution layers, five max pooling layers, two fully-connected layers and one softmax activation layer at the output [50]. The VGG19 model comprises of 16 convolution layers, five max pooling layers, three fully-connected layers and one softmax activation layer at the output [51].

The VGG19 and VGG16 models used in this study were built using the Keras pre-trained VGG library. As with CNN-C1 and CNN-C2, the output layer was configured to have two units with a softmax output representing the probability on an image falling into either classification.

*c) Generative adversarial network (GAN):* GANs are a class of DL algorithms that has been applied to image classification problems [52]. The structure of a GAN, shown in Fig. 4 [53], comprises of two sub-neural networks: a generator network and a discriminator network.

During training, both the generator and the discriminator learn concurrently. The function of a generator network is to produce new, artificial instances of data/images from the input features [52]. This is a form of data augmentation that occurs within the network architecture. The artificial images output from the generator network are evaluated by the discriminator to determine whether they adequately resemble images from the true training dataset. Back-propagation is then used to iteratively train the generator. Generator networks are typically seeded with randomized noise data.

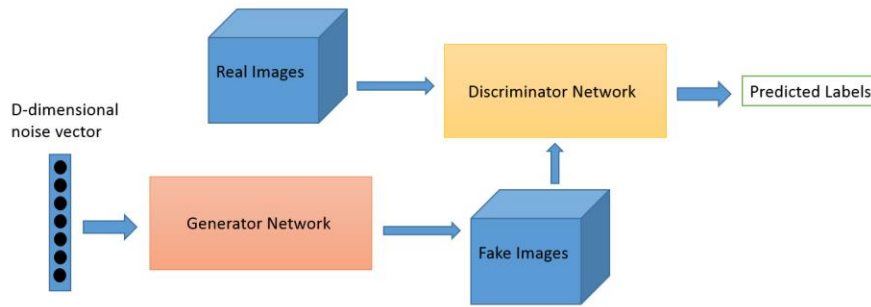


Fig. 4. Structure of a GAN [53]

TABLE IV. ARCHITECTURAL LAYERS OF GAN (DISCRIMINATOR)

Layer Number	Properties
1	Convolutional Layer. 128 nodes, same padding, 3×3 kernel.
2	Max pooling layer
3	Convolutional Layer. 128 nodes, same padding, 3×3 kernel.
4	Max pooling layer
5	Convolutional Layer. 128 nodes, same padding, 3×3 kernel.
6	Max pooling layer
7	Flatten layer
8	Dense Softmax output layer with two units

The discriminator network is trained with images from both the actual dataset and the artificial images produced by the generator. When using a GAN, the discriminator is the final trained model that is tested and deployed in a system.

In the design of a GAN, the discriminator is often a CNN model, and the generator is often a de-convolutional neural network.

The GAN models in this study were built with the architectural layers described in Table IV. There were three convolutional layers used in this network with each layer having 128 nodes. Each convolutional layer was followed by a pooling layer to perform down-sampling. The data was then flattened and passed to a two-unit softmax output layer, where the output prediction was produced. The GAN models deployed a Leaky ReLU activation function, as described by (8), which was added to each down-sampling layer and dense layer. The Leaky ReLU activation function dampens the effect of negative values [54], but does not force them to zero like the standard ReLU function in (7).

$$f_{\text{Leaky ReLU}}(x) = \begin{cases} 0.01x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases} \quad (8)$$

TABLE V. ARCHITECTURAL LAYERS OF MLP

Layer Number	Properties
1	Flatten Layer
2	Dense layer with 128 nodes.
3	Dense Softmax output layer with two nodes.

		True Eye State from Image	
		Open	Closed
Eye State Classification from Model	Open	True Positive	False Positive
	Closed	False Negative	True Negative

Fig. 5. Definitions matrix of model output classifications

d) *Multi-layer perceptron (MLP)*: The MLP is a more basic DL architecture than those derived from the CNN, as it only consists of fully-connected layers [55, 56]. The typical structure of an MLP consists of an input layer, an output layer and at least one hidden layer between the input and output layers. As such, the operation of the MLP is the same as classification stage of a CNN. As a result, MLPs require data to be flattened at the input layer.

The MLP models in this study were built according to the architectural layers described in Table V. The ReLU activation function was implemented in the hidden layer.

4) *Model evaluation*: When analysing model performance, this study considers a range of metrics collectively to provide a holistic evaluation of performance. The following performance metrics were used to evaluate the DL models: accuracy score, precision, negative class precision, recall, specificity and F1-score. These metrics are defined in (9) – (14), in terms of the number of true positive classifications ( $T^+$ ), the number of true negative classifications ( $T^-$ ), the number of false positive classifications ( $F^+$ ) and the number of false negative classifications ( $F^-$ ). These output classifications relate true eye state (based on the known label associated with an image) to the detected eye state (based on the output of the model). The definitions of the different output classifications are visually represented in Fig. 5.

a) *Accuracy score*: The accuracy score is a measure of how many correct predictions were made by the classifier, out of all the predictions made [57, 58]. This is hence the percentage of true output classifications with respect to all output classifications, and is mathematically described as:

$$\text{Accuracy Score} = \frac{T^+ + T^-}{T^+ + T^- + F^+ + F^-} \quad (9)$$

b) *Recall and specificity*: Recall defines how well the model can correctly classify positive outcomes [58, 59]. In the context of this study, recall indicates how many images of open eyes were correctly classified by the model. In addition, for a balanced evaluation of the predictions made for both class labels, the specificity metric was also used. Specificity indicates how well the model can correctly classify negative outcomes [58]. In the context of this study, it indicates how many images of closed eyes were correctly classified by the model. For the problem of driver drowsiness detection, being able to correctly identify when the driver's eyes are closed is of equal importance than identifying when the eye state is open. The mathematical definitions of recall and specificity are given in (10) and (11), respectively.

$$\text{Recall} = \frac{T^+}{T^+ + F^-} \quad (10)$$

$$\text{Specificity} = \frac{T^-}{T^- + F^+} \quad (11)$$

c) *Precision*: Precision defines how well a model can make classify positive outputs [60]. In the context of this study, this indicates the percentage of correct open eye state classifications from all open eye state classifications, as shown by (12). Similarly, the negative class precision represents the percentage of correct closed eye state classifications from all closed eye state classifications. The formula for negative class precisions is presented in (13).

$$\text{Precision} = \frac{T^+}{T^+ + F^+} \quad (12)$$

$$\text{Neg. Class Precision} = \frac{T^-}{T^- + F^-} \quad (13)$$

d) *F1-Score*: The F1-Score represents a weighted average between precision and recall and is hence considered the most appropriate measure of model performance in some literature [57, 61]. Equation (14) presents the mathematical formula to calculate F1-Score [61, 62].

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

#### IV. RESULTS AND DISCUSSION

This section presents and analyses the effects of data augmentation on model performance. Pre-treatment and post-treatment results are presented in Table VI and Table VII, and their descriptive statistics are presented in Table VIII. The

change in performance metrics due to treatment is presented in Table IX. While results for all performance metrics are presented, the main analysis focuses mostly on accuracy and F1-score, as the latter provides insight into the underlying precision and recall.

In the analysis carried out, the authors first confirmed that the DA techniques adopted in this study have improved the performance of the DL models that were investigated. Fig. 6 presents a box-and-whisker diagram of the statistical distribution of all evaluation metrics considered; and compares pre-treatment results with post-treatment results. From the results in Fig. 6, Table VII, Table VIII and Table IX, the following observations and interpretations were made:

1) The post-treatment mean and median values of all evaluation metrics are higher than the pre-treatment values (Table VIII and Table IX). This indicates that the average performance of all DL models studied improved due to the DA techniques applied. The average improvement of the most conclusive metrics, accuracy and F1-score, were +6.1% and +6.8% respectively.

2) The interquartile ranges (IQRs) and standard deviations of post-treatment results were less than for pre-treatment results. In terms of the most conclusive metrics, accuracy and F1-Score, the IQR of both metrics decreased from 13% to 3%. The standard deviation of accuracy scores decreased from 0.17 to 0.12. Similarly, the standard deviation of F1-Scores decreased from 0.20 to 0.14. This indicates that there is less variability in the expected post-treatment performance of all DL algorithms.

3) Outliers were noted in the results, which are clearly illustrated in Fig. 6. These arose from the VGG16 and VGG19 models which were trained on Dataset 1 and displayed inferior performance to the other models studied. Upon investigation, this has been attributed to the dimensionality mismatch between Dataset 1 images (96×96 pixels) and the input dimensions defined by the VGG16 and VGG19 architectures (224×224). While the application of DA techniques has shown the greatest improvement to these models, the post-treatment performance is still low compared to the other models studied. It is thus concluded that the VGG models are not suitable for Dataset 1, and in practice, should not be used with low-resolution cameras that produce smaller video frames/images.

TABLE VI. PRE-TREATMENT AND POST-TREATMENT EVALUATION METRICS (ACCURACY, RECALL AND SPECIFICITY)

Algorithm	Dataset	Accuracy Score		Recall		Specificity	
		Pre-treatment	Post-treatment	Pre-treatment	Post-treatment	Pre-treatment	Post-treatment
CNN-C1	1	93.8%	97.6%	100%	100%	88%	96%
	2	96.5%	97.9%	94%	99%	99%	97%
VGG19	1	47.6%	60.8%	24%	28%	71%	93%
	2	82.4%	95.8%	65%	94%	100%	98%
VGG16	1	50.4%	67.8%	10%	44%	91%	91%
	2	93.1%	95.2%	88%	92%	98%	99%
CNN-C2	1	91.4%	96.2%	100%	97%	83%	97%
	2	95.9%	96.9%	93%	94%	99%	100%
GAN	1	95.4%	97.0%	100%	100%	91%	94%
	2	96.2%	97.2%	97%	97%	95%	98%
MLP	1	91.8%	93.8%	100%	93%	84%	94%
	2	82.1%	93.4%	97%	90%	67%	97%

TABLE VII. PRE-TREATMENT AND POST-TREATMENT EVALUATION METRICS (PRECISION, NEGATIVE CLASS PRECISION AND F1-SCORE)

Algorithm	Dataset	Precision		Neg. Class Precision		F1-Score	
		Pre-treatment	Post-treatment	Pre-treatment	Post-treatment	Pre-treatment	Post-treatment
CNN-C1	1	89%	96%	100%	100%	94%	98%
	2	99%	97%	95%	99%	97%	98%
VGG19	1	45%	81%	48%	57%	31%	61%
	2	100%	98%	74%	94%	82%	96%
VGG16	1	52%	83%	50%	62%	50%	58%
	2	98%	99%	89%	92%	93%	95%
CNN-C2	1	85%	96%	100%	97%	91%	96%
	2	99%	100%	93%	94%	96%	97%
GAN	1	92%	94%	100%	100%	95%	97%
	2	95%	98%	97%	97%	96%	97%
MLP	1	86%	94%	100%	93%	92%	94%
	2	75%	96%	96%	91%	82%	93%

TABLE VIII. DESCRIPTIVE STATISTICS OF EVALUATION METRICS

		Accuracy Score	Recall	Specificity	Precision	Neg. Class Precision	F1-Score
<b>Mean</b>	Pre-Treatment	85%	81%	89%	85%	87%	83%
	Post-Treatment	91%	86%	96%	94%	90%	90%
<b>First Quartile</b>	Pre-Treatment	82%	82%	84%	83%	85%	82%
	Post-Treatment	94%	92%	94%	94%	92%	94%
<b>Median</b>	Pre-Treatment	92%	96%	91%	91%	96%	93%
	Post-Treatment	96%	94%	97%	96%	94%	96%
<b>Third Quartile</b>	Pre-Treatment	96%	100%	98%	98%	100%	95%
	Post-Treatment	97%	98%	98%	98%	98%	97%
<b>IQR</b>	Pre-Treatment	13%	18%	15%	16%	15%	13%
	Post-Treatment	3%	6%	4%	4%	6%	3%
<b>Standard Deviation</b>	Pre-Treatment	0.17	0.30	0.10	0.18	0.18	0.20
	Post-Treatment	0.12	0.23	0.03	0.06	0.14	0.14

TABLE IX. EFFECT OF TREATMENT ON EVALUATION METRICS

Algorithm	Dataset	Accuracy Change	Recall Change	Specificity Change	Precision Change	Neg. Class Precision Change	F1-Score Change
<b>CNN-C1</b>	1	+3.8%	0%	+8%	+7%	0%	+4%
	2	+1.4%	+5%	-2%	-2%	+4%	+1%
<b>VGG19</b>	1	+13.2%	+4%	+22%	+36%	+9%	+30%
	2	+13.4%	+29%	-2%	-2%	+20%	+14%
<b>VGG16</b>	1	+17.4%	+34%	0%	+31%	+12%	+8%
	2	+2.1%	+4%	+1%	+1%	+3%	+2%
<b>CNN-C2</b>	1	+4.8%	-3%	+14%	+11%	-3%	+5%
	2	+1.0%	+1%	+1%	+1%	+1%	+1%
<b>GAN</b>	1	+1.6%	0%	+3%	+2%	0%	+2%
	2	+1.0%	0%	+3%	+3%	0%	+1%
<b>MLP</b>	1	+2.0%	-7%	+10%	+8%	-7%	+2%
	2	+11.4%	-7%	+30%	+21%	-5%	+11%
<b>Average</b>		<b>+6.1%</b>	<b>+5.0%</b>	<b>+7.3%</b>	<b>+9.8%</b>	<b>+2.8%</b>	<b>+6.8%</b>



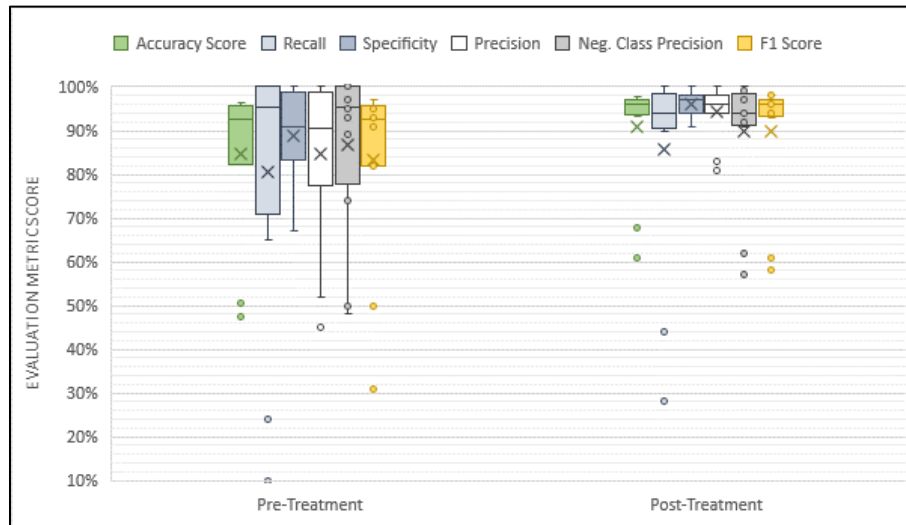


Fig. 6. Statistical distribution of pre-treatment and post-treatment evaluation metrics

TABLE X. EFFECT OF TREATMENT ON EVALUATION METRICS – OUTLIERS REMOVED

Algorithm	Dataset	Accuracy Change	Recall Change	Specificity Change	Precision Change	Neg. Class Precision Change	F1 Score Change
CNN-C1	1	+3.8%	0%	+8%	+7%	0%	+4%
	2	+1.4%	+5%	-2%	-2%	+4%	+1%
VGG19	2	+13.4%	+29%	-2%	-2%	+20%	+14%
VGG16	2	+2.1%	+4%	+1%	+1%	+3%	+2%
CNN-C2	1	+4.8%	-3%	+14%	+11%	-3%	+5%
	2	+1.0%	+1%	+1%	+1%	+1%	+1%
GAN	1	+1.6%	0%	+3%	+2%	0%	+2%
	2	+1.0%	0%	+3%	+3%	0%	+1%
MLP	1	+2.0%	-7%	+10%	+8%	-7%	+2%
	2	+11.4%	-7%	+30%	+21%	-5%	+11%
<b>Average</b>		<b>+4.3%</b>	<b>+2.2%</b>	<b>+6.6%</b>	<b>+5.0%</b>	<b>+1.3%</b>	<b>+4.3%</b>

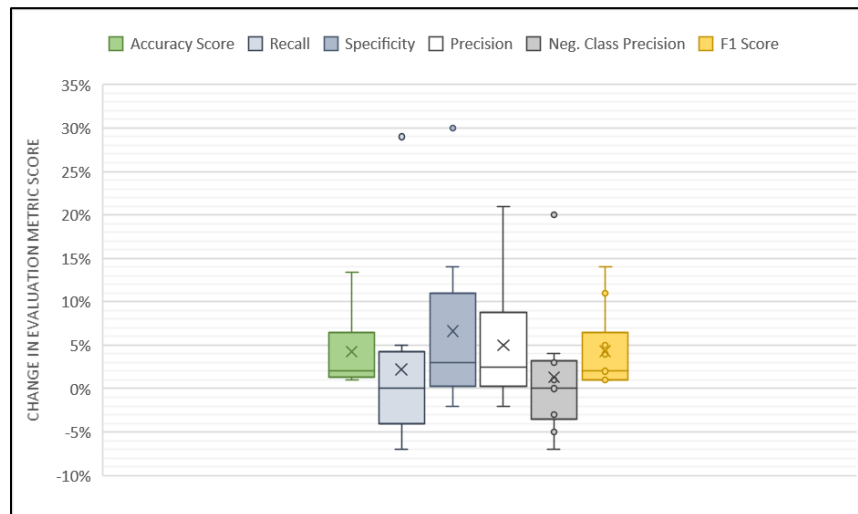


Fig. 7. Statistical distribution of change in evaluation metric scores

Having confirmed the hypothesis that the DA techniques that were applied have improved the performance of the DL models studied, the next step was to attempt to quantify the extent of this improvement. The VGG16 and VGG19 models trained on Dataset 1 were excluded from this analysis due to their poor performance, as discussed previously. Table X presents the change in evaluation metrics due to the application of DA with these models removed. The statistical distribution of the data presented in Table X is illustrated in Fig. 7.

When analyzing the results, the following was observed:

1) A few instances were observed where applying DA treatment caused a reduction in individual evaluation metrics (recall, precision, specificity and negative class precision), as indicated by shaded backgrounds within Table X. However, despite this, the F1-Score increased for all models, indicating that these performance reductions were compensated for. The average increase in both accuracy and F1-Score was +4.3%, and the median increase in each of these metrics were +2.1% (accuracy) and +2.0% (F1-Score).

2) The box-and-whisker diagrams in Fig. 7 indicated that there is significantly more variability for recall, specificity, precision and negative-class precision than for accuracy and F1-Scores. As such, attempts at quantifying the expected improvement in DL model performance using the methods in this study can only reasonable be performed for accuracy and F1-Score. However, these are the most conclusive metrics to evaluate the DL models studied.

3) By analyzing the distribution of the change in accuracy and F1-Scores, it was observed that the data for these evaluation metrics was positively skewed. This resulted from the high pre-treatment accuracy scores and F1-Scores of some of the DL models studied, where there was not much room for improvement without over-fitting the model to the training dataset.

Prompted by the final observation listed above, the final analysis investigated the relationship between the change in evaluation metric scores and pre-treatment metric scores. The scatterplot presented in Fig. 8 illustrates this relationship, using data from Table VI, Table VII and Table X and excluding the outlier results resulting from the VGG16 and VGG19 models that were trained on Dataset 1. The trend lines show that all evaluation metrics exhibited a strong negative correlation, indicated by the  $R^2$  values of the correlation trend lines ( $R^2 > 0.7$  for all evaluation metrics). From this, it is concluded that the DA techniques under study have a marginal improvement when applied to DL models that already exhibit strong performance, but are much more powerful in enhancing weaker-performing DL models. From Fig. 8, an improvement of  $\geq +5\%$  to an evaluation metric occurs when the pre-treatment value of the metric is  $\leq 90\%$ . This indicates the type of DL models for driver drowsiness detection that will benefit most from the DA techniques presented in this study, and is recommended to developers as a design guideline when considering the implementation of the DA techniques presented in this paper.

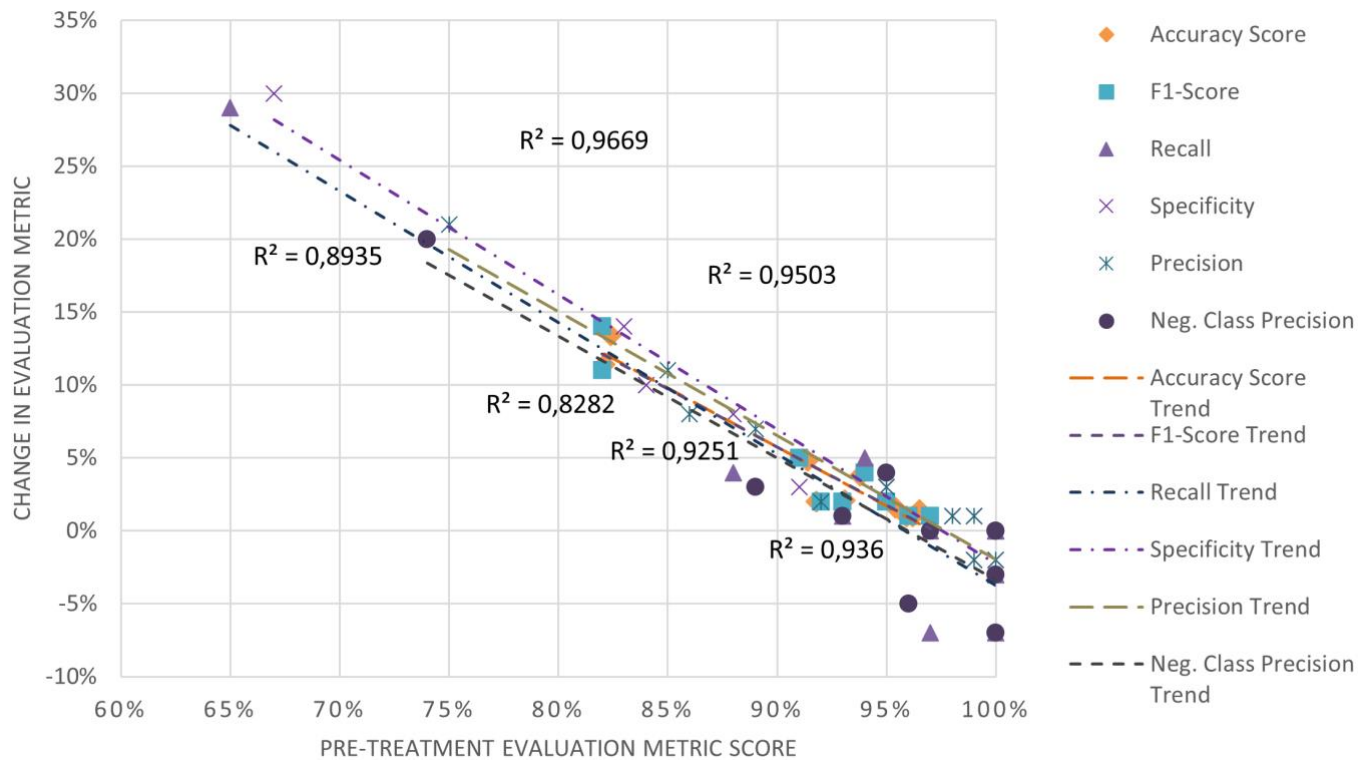


Fig. 8. Relationship between pre-treatment evaluation metric score and change in evaluation metric score

The results confirm that by modelling real-world scenarios using the data augmentation techniques described in Section II.B.2, it is possible to train more robust deep learning algorithms that perform driver drowsiness detection. With respect to implementation of driver drowsiness detection systems, the deep learning model development and training would be performed before the model is deployed in the driver drowsiness detection system hardware.

## V. CONCLUSION

Many road accidents are caused by driver drowsiness. Previous studies have considered applying deep learning techniques to detect driver drowsiness and improve road traffic safety. In practically testing their systems, many previous studies have indicated that real-world scenarios such as unfavourable ambient lighting and movement of the driver while driving cause inaccuracies when detecting driver drowsiness.

In this study, the authors focussed on the deep learning algorithms that determine driver drowsiness based on the eye state of the driver. It was hypothesised that by modelling the real-world scenarios and using data augmentation techniques on a standardised image dataset, the performance of the DL models would improve. This study considered two different datasets, six different DL models: two CNN variations (CNN-C1 and CNN-C2), two architectures designed by the VGG (VGG16 and VGG19), a GAN and an MLP.

The performance of the DL models was evaluated primarily using accuracy and F1-Score, although other metrics such as precision, recall, specificity and negative class precision were also considered. In analyzing the results in aggregation, improvements across all metrics were noted. The average improvement in accuracy across all DL models was +6.1% and the average improvement in F1-Score was 6.8%, and the variability in model performance was reduced. However, there were some challenges noted when training the VGG models. These models trained on low-resolution images, exhibited poor performance and distorted these results. A more realistic indication of the benefits of DA for the DL models studied was obtained by excluding these outliers, yielding an average improvement of +4.3% for both accuracy and F1-Score.

The results further indicated that the extent to which the DA techniques studied improve DL model performance is strongly correlated with the pre-treatment DL model performance. From the analysis conducted, the data augmentation techniques presented are best suited for improving models with accuracy and F1-Scores  $\leq 90\%$  - although they are applicable to any DL model for driver drowsiness detection.

It was thus concluded that the use of DA techniques improves the performance of DL models for driver drowsiness detection under the isolated conditions of this study. However, since the conditions of this study focussed on testing the DL models on images from datasets, rather than testing being done on captured data from a real-world driver drowsiness detection system, this opens the possibility for future research. Future works should look at implementing the trained DL models

proposed in this study in practical driver drowsiness detection systems to validate these results.

## DATA AVAILABILITY

The drowsiness detection dataset (Dataset 1) is available online at: <https://www.kaggle.com/prasadvpatil/mrl-dataset>.

The yawn eye dataset (Dataset 2) is available online at: <https://www.kaggle.com/serenaraju/yawn-eye-dataset-new>.

## FUNDING

All funding in support of this research was provided by the Durban University of Technology.

## CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

## ACKNOWLEDGMENT

The authors acknowledge the support of the wider Faculty of Accounting and Informatics at the Durban University of Technology, and Mr Shaldon Wade Naidoo for his assistance.

## REFERENCES

- [1] D. Machetele and K. Yessoufou, "A Decade Long Slowdown in Road Crashes and Inherent Consequences Predicted for South Africa," (in English), *Frontiers in Future Transportation, Original Research* vol. 2, 2021-November-03 2021, doi: 10.3389/ffutr.2021.760640.
- [2] T. Verster and E. Fourie, "The good, the bad and the ugly of South African fatal road accidents," *South African Journal of Science*, vol. 114, pp. 63-69, 2018. [Online]. Available: [http://www.scielo.org.za/scielo.php?script=sci\\_arttext&pid=S0038-23532018000400017&nrm=iso](http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S0038-23532018000400017&nrm=iso).
- [3] M. Mokoatle, V. Marivate, and M. E. Bukohwo, "Predicting Road Traffic Accident Severity using Accident Report Data in South Africa," presented at the Proceedings of the 20th Annual International Conference on Digital Government Research, Dubai, United Arab Emirates, 2019. [Online]. Available: <https://doi.org/10.1145/3325112.3325211>.
- [4] L. Weisberg, "The Long-Term Effects of Car Accident Injuries." Weisberg Cummings, P.C. Employment Law Attorneys <https://www.weisbergcummings.com/long-term-effects-car-accident-injuries/> (accessed 10 November, 2022).
- [5] A. B. Boye, "The Effect of Road Traffic Delay on Supply Chain Performance of Manufacturing Firms in Lagos State, Nigeria," *International Journal of Management Technology*, vol. 5, no. 3, pp. 9-20, 2018. [Online]. Available: <https://www.eajournals.org/journals/international-journal-of-management-technology-ijmt/vol-5-issue-3-september-2018/the-effect-of-road-traffic-delay-on-supply-chain-performance-of-manufacturing-firms-in-lagos-state-nigeria/>.
- [6] MiWay. "The truth about truck accidents on SA roads." <https://www.miway.co.za/blog/business-insurance/the-truth-about-truck-accidents-on-sa-roads> (accessed 10 November, 2022).
- [7] Dovetail. "6 Road Freight Challenges in South Africa." <https://dovetail.co.za/6-road-freight-challenges-in-south-africa/> (accessed 10 November, 2022).
- [8] Daily News Reporter. "SA has the world's poorest road safety records - WHO report." IOL News. <https://www.iol.co.za/the-post/news/sa-has-the-worlds-poorest-road-safety-records-who-report-18631896> (accessed 10 November, 2022).
- [9] L. Rondganger. "South Africa's roads deaths are a 'national crisis'." IOL News. <https://www.iol.co.za/news/south-africa/kwazulu-natal/south-africas-roads-deaths-are-a-national-crisis-cefc54fe-bafe-45c6-b0d9-f7c4b1ce7ea8> (accessed 10 November, 2022).
- [10] B. Alshaqai, A. S. Baquhaizel, M. E. A. Ouis, M. Boumehed, A. Ouamri, and M. Keche, "Driver drowsiness detection system," in 2013

- 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA), 12-15 May 2013 2013, pp. 151-155, doi: 10.1109/WoSSPA.2013.6602353.
- [11] L. J. Hurwitz. "Driver Fatigue: A Leading Cause of Accidents and Death in the Transportation Industry." Segal McCambridge. <https://www.segalmccambridge.com/blog/driver-fatigue-a-leading-cause-of-accidents-and-death-in-the-transportation-industry/> (accessed 10 November, 2022).
- [12] M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, "A Survey on State-of-the-Art Drowsiness Detection Techniques," IEEE Access, vol. 7, pp. 61904-61919, 2019, doi: 10.1109/ACCESS.2019.2914373.
- [13] R. Jabbar, K. Al-Khalifa, M. Kharbeche, W. Alhajyaseen, M. Jafari, and S. Jiang, "Real-time Driver Drowsiness Detection for Android Application Using Deep Neural Networks Techniques," Procedia Computer Science, vol. 130, pp. 400-407, 2018/01/01/ 2018, doi: <https://doi.org/10.1016/j.procs.2018.04.060>.
- [14] R. Jabbar, M. Shinoy, M. Kharbeche, K. Al-Khalifa, M. Krichen, and K. Barkaoui, "Driver Drowsiness Detection Model Using Convolutional Neural Networks Techniques for Android Application," in 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIOT), 2-5 Feb. 2020 2020, pp. 237-242, doi: 10.1109/ICIOT48696.2020.9089484.
- [15] S. S. Sharan, R. Viji, R. Pradeep, and V. Sajith, "Driver Fatigue Detection Based On Eye State Recognition Using Convolutional Neural Network," in 2019 International Conference on Communication and Electronics Systems (ICCES), 17-19 July 2019 2019, pp. 2057-2063, doi: 10.1109/ICCES45898.2019.9002215.
- [16] V. R. R. Chirra, S. R. Uyyala, and V. K. K. Kolli, "Deep CNN: A Machine Learning Approach for Driver Drowsiness Detection Based on Eye State," Revue d'Intelligence Artificielle, vol. 33, no. 6, pp. 461-466, 2019, doi: 10.18280/ria.330609
- [17] M. Hashemi, A. Mirrashid, and A. B. Shirazi, "Driver Safety Development: Real-Time Driver Drowsiness Detection System Based on Convolutional Neural Network," SN Computer Science, vol. 1, no. 5, p. 289, 2020/08/31 2020, doi: 10.1007/s42979-020-00306-9.
- [18] F. Zhang, J. Su, L. Geng, and Z. Xiao, "Driver Fatigue Detection Based on Eye State Recognition," in 2017 International Conference on Machine Vision and Information Technology (CMVIT), 17-19 Feb. 2017 2017, pp. 105-110, doi: 10.1109/CMVIT.2017.25.
- [19] W. Kongcharoen, S. Nuchitprasitchai, Y. Nilsiam, and J. M. Pearce, "Real-Time Eye State Detection System for Driver Drowsiness Using Convolutional Neural Network," in 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 24-27 June 2020 2020, pp. 551-554, doi: 10.1109/ECTI-CON49241.2020.9158265.
- [20] M.-C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, "Multilayer perceptron and neural networks," WSEAS Transactions on Circuits and Systems, vol. 8, no. 7, pp. 579-588, 2009, doi: 10.5555/1639537.1639542.
- [21] T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," in 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), 10-12 March 2017 2017, pp. 721-724, doi: 10.1109/ICBDA.2017.8078730.
- [22] R. Seetharaman, S. Sridhar, and S. Mootha, "Detection and State Analysis of Drowsiness using Multitask Learning with Neural Networks," in 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS), 21-23 Oct. 2020 2020, pp. 1-8, doi: 10.1109/ICDS50568.2020.9268740.
- [23] T. Vesselenyi, S. Moca, A. Rus, T. Mitran, and B. Tătaru, "Driver drowsiness detection using ANN image processing," IOP Conference Series: Materials Science and Engineering, vol. 252, no. 1, p. 012097, 2017/10/01 2017, doi: 10.1088/1757-899X/252/1/012097.
- [24] A. Ghourabi, H. Ghazouani, and W. Barhoumi, "Driver Drowsiness Detection Based on Joint Monitoring of Yawning, Blinking and Nodding," in 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP), 3-5 Sept. 2020 2020, pp. 407-414, doi: 10.1109/ICCP51029.2020.9266160.
- [25] P. Bajaj, R. Ray, S. Shedje, S. Jaikar, and P. More, "Synchronous System for Driver Drowsiness Detection Using Convolutional Neural Network, Computer Vision and Android Technology," in 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 19-20 March 2021 2021, vol. 1, pp. 340-346, doi: 10.1109/ICACCS51430.2021.9441670.
- [26] H. Ahuja, S. Saurav, S. Srivastava, and C. Shekhar, "Driver Drowsiness Detection using Knowledge Distillation Technique for Real Time Scenarios," in 2020 IEEE 17th India Council International Conference (INDICON), 10-13 Dec. 2020 2020, pp. 1-5, doi: 10.1109/INDICON49873.2020.9342263.
- [27] M. Ngxande, J. R. Tapamo, and M. Burke, "Bias Remediation in Driver Drowsiness Detection Systems Using Generative Adversarial Networks," IEEE Access, vol. 8, pp. 55592-55601, 2020, doi: 10.1109/ACCESS.2020.2981912.
- [28] P. Chen, S. Liu, H. Zhao, and J. Jia, "Gridmask data augmentation," arXiv preprint arXiv:2001.04086, 2020.
- [29] H. J. Weerts, A. C. Mueller, and J. Vanschoren, "Importance of tuning hyperparameters of machine learning algorithms," arXiv preprint arXiv:2007.07588, 2020.
- [30] P. V. Patil. Drowsiness Detection Dataset, Kaggle. [Online]. Available: <https://www.kaggle.com/prasadvpatil/mrl-dataset>
- [31] S. Raju. yawn\_eye\_dataset\_new, Kaggle. [Online]. Available: <https://www.kaggle.com/serenaraju/yawn-eye-dataset-new>
- [32] A. Takimoglu. "What Is Data Augmentation? Techniques, Benefit & Examples." AIMultiple. <https://research.aimultiple.com/data-augmentation/> (accessed 8 November, 2022).
- [33] N. Ketkar, "Introduction to Keras," in Deep Learning with Python: Springer, 2017, pp. 97-111.
- [34] A. Bhandari. "Image Augmentation on the Fly Using Keras Imagedatagenerator." Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2020/08/image-augmentation-on-the-fly-using-keras-imagedatagenerator/> (accessed).
- [35] F. Chollet. "Building powerful image classification models using very little data." The Keras Blog. <https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html> (accessed 8 November, 2022).
- [36] J. Brownlee. "How to Normalize, Center, and Standardize Image Pixels in Keras." Machine Learning Mastery. <https://machinelearningmastery.com/how-to-normalize-center-and-standardize-images-with-the-imagedatagenerator-in-keras/> (accessed 10 November, 2022).
- [37] J. Brownlee. "A Gentle Introduction to Batch Normalization for Deep Neural Networks." Machine Learning Mastery. <https://machinelearningmastery.com/batch-normalization-for-training-of-deep-neural-networks/2020> (accessed 10 November, 2022).
- [38] M. Alam. "Data Normalization in Machine Learning." Towards Data Science. <https://towardsdatascience.com/data-normalization-in-machine-learning-395fdec69d02> (accessed 7 November, 2022).
- [39] TheAILEarner. "Data Augmentation with Keras Imagedatagenerator." TheAILEarner. <https://theailearner.com/2019/07/06/data-augmentation-with-keras-imagedatagenerator/> (accessed 7 November, 2022).
- [40] S. Albawi, T. A. Mohammed, and S. Al-Azawi, "Understanding of a convolutional neural network," in 2017 International Conference on Engineering and Technology (ICET), 21-23 Aug. 2017 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.
- [41] M. K. Gurucharan. "Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network." upGrad Education Private Limited. <https://www.upgrad.com/blog/basic-cnn-architecture/> (accessed 1 December, 2021).
- [42] Y. Wang, Y. Li, Y. Song, and X. Rong, "The Influence of the Activation Function in a Convolution Neural Network Model of Facial Expression Recognition," Applied Sciences, vol. 10, no. 5, p. 1897, 2020. [Online]. Available: <https://www.mdpi.com/2076-3417/10/5/1897>.
- [43] F. Chollet. "The Sequential Model." Keras.io. [https://keras.io/guides/sequential\\_model/](https://keras.io/guides/sequential_model/) (accessed 8 November, 2022).
- [44] R. Keshari, M. Vatsa, R. Singh, and A. Noore, "Learning Structure and Strength of CNN Filters for Small Sample Size Training," in 2018

- IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18-23 June 2018 2018, pp. 9349-9358, doi: 10.1109/CVPR.2018.00974.
- [45] A. Wiranata, S. A. Wibowo, R. Patmasari, R. Rahmania, and R. Mayasari, "Investigation of Padding Schemes for Faster R-CNN on Vehicle Detection," in 2018 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), 5-7 Dec. 2018 2018, pp. 208-212, doi: 10.1109/ICCEREC.2018.8712086.
- [46] Z. J. Wang et al., "CNN 101: Interactive visual learning for convolutional neural networks," in Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1-7.
- [47] G. Dumane. "Introduction to Convolutional Neural Network (CNN) Using Tensorflow." Towards Data Science. <https://towardsdatascience.com/introduction-to-convolutional-neural-network-cnn-de73f69c5b83> (accessed 8 November, 2022).
- [48] 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.
- [49] OpenGenus IQ. "Understanding the VGG19 Architecture." OpenGenus IQ: Computing Expertise & Legacy. <https://iq.opengenus.org/vgg19-architecture/> (accessed 31 October, 2021).
- [50] T. J. Perumanoor. "What Is VGG16? — Introduction to VGG16." Great Learning. <https://medium.com/@mygreatlearning/what-is-vgg16-introduction-to-vgg16-f2d63849f615> (accessed 17 November, 2021).
- [51] L. Wen, X. Li, X. Li, and L. Gao, "A New Transfer Learning Based on VGG-19 Network for Fault Diagnosis," in 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD), 6-8 May 2019, pp. 205-209, doi: 10.1109/CSCWD.2019.8791884.
- [52] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 53-65, 2018, doi: 10.1109/MSP.2017.2765202.
- [53] C. Nicholson. "A Beginners Guide to Generative Adversarial Networks (GANs)." Pathmind. <https://wiki.pathmind.com/generative-adversarial-network-gan> (accessed 4 March, 2022).
- [54] A. K. Dubey and V. Jain, "Comparative Study of Convolution Neural Network's Relu and Leaky-Relu Activation Functions," in Applications of Computing, Automation and Wireless Systems in Electrical Engineering, Singapore, S. Mishra, Y. R. Sood, and A. Tomar, Eds., 2019// 2019: Springer Singapore, pp. 873-880.
- [55] H. Zhu, D. Gao, and S. Zhang, "A Perceptron Algorithm for Forest Fire Prediction Based on Wireless Sensor Networks," Journal on Internet of Things, vol. 1, no. 1, 2019, doi: 10.32604/jiot.2019.05897.
- [56] J. Brownlee. "Perceptron Algorithm for Classification in Python." Towards Data Science. <https://machinelearningmastery.com/perceptron-algorithm-for-classification-in-python/> (accessed 4 April, 2022).
- [57] D. R. Govinda and Z. Waheed, "Increasing Trend in Accuracy Score for Machine Learning Algorithms," International Journal of Development and Public Policy, vol. 1, no. 5, pp. 180-182, 10/23 2021. [Online]. Available: <https://www.openaccessjournals.eu/index.php/ijdpp/article/view/429>.
- [58] S. Raschka and V. Mirjalili, Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2. Packt Publishing Ltd, 2019.
- [59] K. Roth, T. Milbich, S. Sinha, P. Gupta, B. Ommer, and J. P. Cohen, "Revisiting Training Strategies and Generalization Performance in Deep Metric Learning," presented at the Proceedings of the 37th International Conference on Machine Learning, Proceedings of Machine Learning Research, 2020. [Online]. Available: <https://proceedings.mlr.press/v119/roth20a.html>.
- [60] M. Dua, Shakshi, R. Singla, S. Raj, and A. Jangra, "Deep CNN models-based ensemble approach to driver drowsiness detection," Neural Computing and Applications, vol. 33, no. 8, pp. 3155-3168, 2021/04/01 2021, doi: 10.1007/s00521-020-05209-7.
- [61] S. Panda and M. Kolhekar, "Feature Selection for Driver Drowsiness Detection," in Proceedings of International Conference on Computational Intelligence and Data Engineering, Singapore, N. Chaki, N. Devarakonda, A. Sarkar, and N. C. Debnath, Eds., 2019// 2019: Springer Singapore, pp. 127-140.
- [62] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, no. 1, p. 6, 2020/01/02 2020, doi: 10.1186/s12864-019-6413-7.

# An Improved Ant Colony Algorithm for Virtual Resource Scheduling in Cloud Computing

## Methods to Improve the Performance of Virtual Resource Scheduling

Chunlei Zhong\*<sup>1</sup>, Gang Yang<sup>2</sup>

Huai'an Bioengineering Branch Institute, Jiangsu Union Technical Institute,  
Huai'an 223200, China<sup>1</sup>  
College of Teacher Education, Wenzhou University,  
Wenzhou 325035, China<sup>2</sup>

**Abstract**—In order to solve the problems of uneven spatial distribution of data nodes and unclear weight relationship of virtual scheduling features in cloud computing platform, a virtual resource scheduling method based on improved ant colony algorithm is studied and designed to improve the performance of virtual resource scheduling in cloud computing platform by this method. After analyzing the information resource sequence change of the cloud computing platform, according to the STR - Tree partition graph, a simulated annealing-based algorithm is employed to classify the resource types after optimal scheduling into IO types, middle types and CPU types, and the time span and load balance are set as the measurement indexes. The simulation results show that after applying this method, the occupied resources of the main platform are 535 MB, which are much lower than the other two comparison algorithms, and the method has improved the allocation rationality, resource balance, maximum queue length and energy consumption. This result indicates that applying this virtual resource scheduling method can effectively improve the intelligent scheduling of virtual resources in the cloud computing platform.

**Keywords**—Improved ant colony algorithm; cloud computing; virtual resources; intelligent scheduling

### I. INTRODUCTION

Cloud computing is the mainstream way to provide services to users relying on resources deployed over numerous physical machines to be available regardless of their location. *Virtualization* is the key concept employed to provide a logical abstraction over physical resources running on data centers. This model considers data centers as nodes suitable for cloud computing and provides a unified and easy-to-use resource organization.

Different users request access and obtain resources according to their own needs, without paying attention to the specific location of resources. In order to provide users with the same or similar functions as the real layer, and to facilitate the stable operation of the upper layer system in the middle layer, cloud virtual resource scheduling is needed. Unlike other distributed computing platforms, virtualization provides more computing power and flexibility for applications, an advantage that fits well with the needs and main concepts of cloud computing [1].

In order to improve the usage of virtual resource in cloud computing, which considers hosting, network traffic, and other costs optimization, is not uncommon for multiple system resources, even from different organizations, to share a single physical server. Besides, a service application can also run over multiple service provider platforms to be accessed by multiple clients simultaneously. This scenario is tackled by the multi-tenant scheduling process, which aims for high resource utilization using the cloud computing network.

Laith & Ali [2] presented a hybrid ant colony optimization algorithm based on elitist differential evolution is proposed to solve the multi-objective task scheduling problem in cloud computing environment, which minimizes the maximum completion time, maximizes the resource utilization, and uses elitist differential evolution as a local search technique to improve its utilization ability and avoid falling into local optimization. In [3], by satisfying the constraint of security requirements, we find an optimal scheduling algorithm with minimum completion time and cost, and users can submit their security requirements to the cloud provider during the negotiation period. However, none of the methods mentioned before has solved the problems of indefinite weight relation of virtual scheduling features and uneven distribution of cloud platform information space, so it is difficult to realize virtual resource scheduling efficiently.

In order to solve the problems of the above methods, a cloud computing virtual resource scheduling method based on improved ant colony algorithm is proposed. The overall research technical route of this method is as follows:

- 1) This paper analyzes the characteristics of the sequence change of information resources on the cloud computing platform, and establishes the index relationship between spatial information objects through STR tree partition graph to solve the problem of data imbalance.
- 2) Simulated annealing algorithm is used to divide the optimized resource types into IO type, intermediate type and CPU type, and time span and load balance are used as indicators to complete the scheduling of virtual resources.
- 3) Experimental verification, taking the occupied space, load fairness, Allocation Rationality, resource balance, the longest queue length and energy consumption value as the test

indicators, the comparative tests of different methods are carried out.

By updating the above techniques, the proposed cloud computing virtual resource scheduling method can effectively solve the problems of uncertain feature weight relationship and uneven information space distribution in the virtual scheduling process, and it is expected that this method will improve the performance of cloud computing virtual resource scheduling and enhance new ideas for the field of virtual resource scheduling.

The structure of the study is divided into four main areas as follows. The first part introduces the current research status of cloud computing resource scheduling and compares it with the proposed algorithm. The second part introduces the virtual resource scheduling method based on the improved ant colony algorithm and describes its practical computing process. The third part is a performance test and application analysis of the virtual resource scheduling algorithm proposed in the study, and the results are organized and analyzed. The last part is a summary and discussion of the whole paper.

## II. CLOUD COMPUTING PLATFORM VIRTUALIZATION RESOURCE EXTRACTION

### A. Changes in the Sequence of Communication Network Information Resources of Cloud Computing Platforms

Each node in the cloud computing platform sends information to the center node to report according to the system load and user's demand [4]. In most cases, passive real-time monitoring is used to schedule resources. The client connects to the cloud server based on the control server.

When the client sends the computer application, Computer application is to study the theory, method, technology and system of computer application in various fields. Computer application is an activity that gives guidelines on how to participate and implement in social activities. The cloud server can store some data related to the user, Distributed storage is used to consolidate a large number of servers into a

supercomputer, providing a large amount of data storage and processing services. Distributed file systems and distributed databases allow access to common storage resources and realize IO sharing of application data files. By integrating computer, the resource scheduling server sends the calculation results to the user's computer, and the computer client returns the results. Calculate the data source through the server and wait for the next user request to obtain the communication network model of the cloud computing platform, as shown in Fig. 1.

In Fig. 2, the resource storage module adopts idt72v3680 memory produced by IDT Company. Idt72v3680 is one of the high-density supersynctm II 36 bit series memories idt72v3640 ~ 3690 of IDT company. Its storage structure is  $16384 \times 36$ . This series of CMOS process FIFO (first in first out) chips have great depth. The storage structure of this memory is  $16384 * 36$ , and the storage space can reach 1.125MB [5-8]. The internal circuit diagram of the memory is shown in Fig. 3.

According to Fig. 3, Memory is a memory device in a computer system that stores programs and data. All information in the computer, including input raw data, computer programs, intermediate running results and final running results are stored in the memory. It stores and retrieves information according to the location specified by the controller. The main function of the memory is to store programs and various data, and to complete the program or data access at high speed and automatically during the running of the computer. The memory has an independent chip structure. The storage rate of the memory can reach 100MHz, the working voltage is 3.15-3.45v, and the maximum working current is 400mA. In addition, the memory uses CMOS technology, belongs to high-density memory, and has great depth. During operation, idt72v3680 memory, as the transmission link between acquisition module and control module, can still realize efficient data storage without additional coordination under the condition of different rates of the two sides.

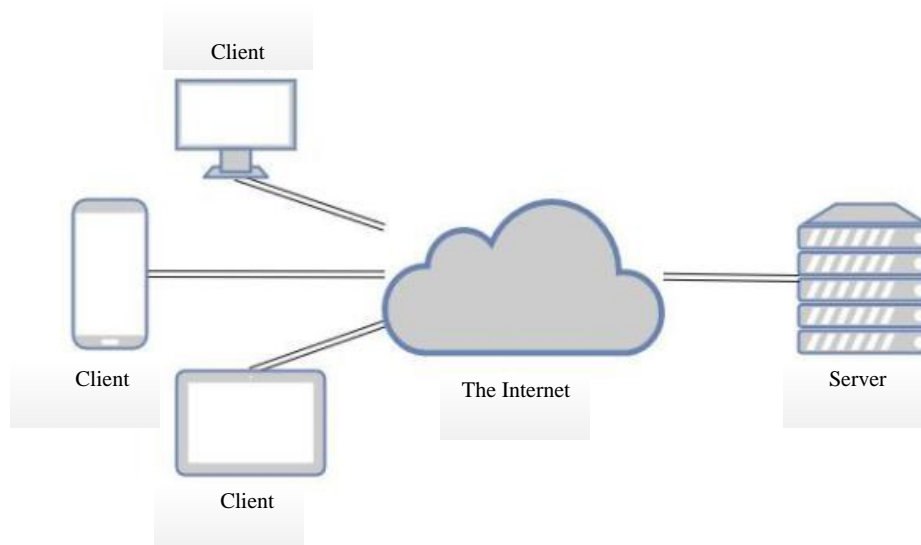


Fig. 1. The cloud server

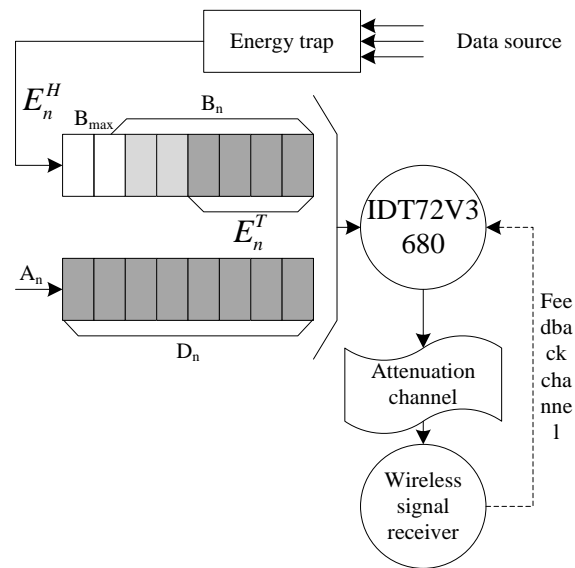


Fig. 2. Communication network model of cloud computing platform

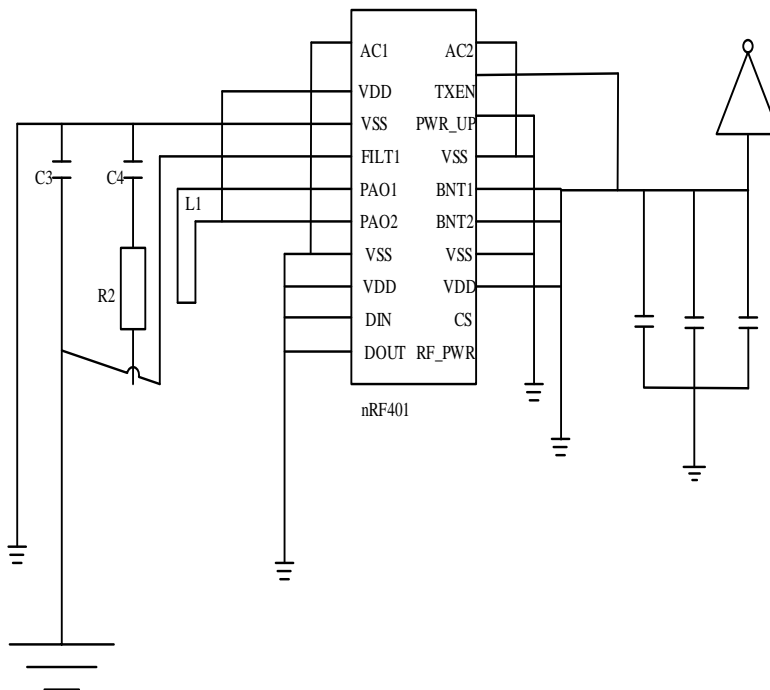


Fig. 3. Internal circuit diagram of memory

If it is simply stored, it is difficult to complete the subsequent indexing steps. Therefore, in the process of establishing the storage mechanism, it is necessary to consider the indexing efficiency and divide and encode the data attributes [9]. The types and sizes of cloud data are different, and it is difficult to divide them. It is necessary to develop a reasonable strategy to divide them effectively [10]. In the process of partition coding, only the size, type and spatial distribution of data need to be considered to ensure the relative integrity of the cloud platform spatial information entity as far as possible, so as not to divide it into multiple scattered data blocks. If two data blocks are adjacent data in the information space, they can be consciously divided into the same interval

and given adjacent codes. In addition, if the data volume difference between the divided intervals is too large, it is easy to lead to the low balance of the distributed cluster, resulting in logical vulnerabilities. Therefore, when coding data nodes in partitions, this kind of data imbalance can be solved through STR tree. The main operation mechanism of STR tree is to establish the index relationship between spatial information objects through the external matrix between two data nodes, and code and sort their abscissa and ordinate between the two dimensions.

In the data storage space of a cloud platform, it is first necessary to determine the number of data partitions  $n$ , code



other data entities into  $A, B, C, \dots, X, Y, Z$ , etc., and divide the average total amount of data block space  $\cdot \sqrt{n}$ . In this way, the horizontal and vertical coordinates of all data can be indexed to  $\frac{N}{\sqrt{n}}$ , all data coordinates can be grouped into different data blocks, and all vertical and horizontal coordinates are  $\frac{N}{\sqrt{n}}$ . Determine the partition coding basis of STR tree, as shown in Fig. 4.

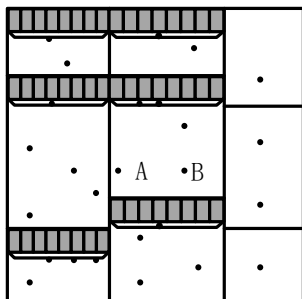


Fig. 4. STR tree zoning diagram

As shown in Fig. 4, STR tree can effectively solve the problem of uneven spatial distribution of information of data nodes in the cloud platform, so as to reduce the repeatability of data nodes. Before designing the virtual resource scheduling model of cloud computing platform, firstly, set the number of cellular users in heterogeneous units as  $m$  and the number of communication network users of cloud computing platform as  $N$ ,  $M = \{1, 2, \dots, M\}$  and  $N = \{1, 2, \dots, N\}$ . Multiple uplink resources are used simultaneously in the scheduling process, and the resource blocks do not affect each other.

In time period  $T$ , the transmission power of the  $n$ th cloud computing platform communication network at the  $m$ -th cellular user and its uplink resources is  $P^m(t)$  and  $P_n^m(t)$  respectively; The signal to dryness ratios of  $N$  on  $M$  to cellular users in time period  $T$  are  $SINR_n^m(t)$  and  $SINR^m(t)$  respectively. Set  $i, j \in N$ . Within the time period  $T$  [11], the communication between the cloud computing platform communication network and the cellular user is dry, as shown in formula (1):

$$SINR_i^m(t) = \frac{P_i^m(t)h_{ii}(t)}{N_0 + \sum_{j \in S_m, i \neq j} P_j^m(t)h_{ji}(t) + P^m(t)h_{mi}(t)} \quad (1)$$

Among them,  $h_{ii}(t), h_{ji}(t), h_{mi}(t)$ , and are the channel power gain of the communication network of the cloud computing platform, which are respectively used to describe the channel power gain of uplink resources from the information transmission end to the information receiving end at different times, and calculate the ratio of output power and input power of uplink resources at different times. If all multiplexing resource sets are  $S_m$  and the noise power spectral density is  $N_0$ , the signal to dryness ratio of the

primary user is:

$$SINR^m(t) = \frac{P^m(t)h_{mB}(t)}{N_0 + \sum_{j \in S_m} P_j^m(t)h_{jB}(t)} \quad (2)$$

Among them, the channel power gains from the user and the network to the base station are  $h_{mB}(t), h_{jB}(t)$  [12]. Calculate the maximum communication rate of cloud computing platform communication network to cellular users through Shannon channel capacity calculation formula:

$$R^m(P^m)_i = \log_2(1 + SINR^m(t)) \quad (3)$$

Where  $P^m$  represents the transmission power passing on the uplink resource.

Only when the uplink interference of the communication network users of the cloud computing platform to the cellular users is minimized can the communication efficiency of the cellular users be maximized. The resource allocation of the resource block can be expressed by formula (4), the time period is  $t$ , and the signal to dryness ratio threshold of the cellular user is  $\zeta_{Threshold}^m$ . Average maximum throughput of communication network resource scheduling in  $t$  [13]:

$$\arg \max = \frac{1}{T} \left[ \sum_{t=1}^T \left( \sum_{m=1}^N \log_2(1 + SINR^m(t)) \right) \right] \quad (4)$$

Among them,  $SINR^m(t) \geq \zeta_{Threshold}^m$ . There are network information resources arriving at the wireless signal transmitter in each time slot, and  $X_n$  is used to represent the number of information resources contained in the communication network, so the change process of the information resource sequence of the communication network can be described as

$$D_{n+1} = \max \{D_n - X_n, 0\} + A_n \quad (5)$$

In formula (1),  $D_{n+1}$  and  $D_n$  respectively represent the information resource sequence  $n+1$  of the communication network of the  $n$ th and  $t$ th time slots;  $A_n$  represents the number of data packets arriving at the wireless signal transmitter in the time slot.

### B. Cloud Platform Virtualization Resource Association Feature Extraction

Advantages of ant colony algorithm over other swarm intelligence algorithms: (1) the algorithm has general applicability: as long as the problem can be transformed into a connected graph, it can generally be solved by ant colony algorithm in scenes with low real-time requirements. (2) Randomness and certainty are interdependent: the heuristic function represents a priori knowledge, its weight represents the degree of certainty in selection, and the pheromone update criterion represents the mutual restriction between the two. Pheromone update, as the performance mechanism of positive feedback, reflects the degree to which the path is focused (certainty), while the volatilization coefficient participates as the negative feedback mechanism, the excellent path with too

much pheromone behind has the opportunity to catch up. The roulette algorithm increases the possibility of exploring a larger solution space (randomness). Under normal circumstances, the cloud platform communication process acts on the attenuation channel, which means that the network channel gain remains the same in any time slot, but there are some differences in the network channel states in different time slots. If the wireless network is in the form of adaptive coding modulation, the calculation formula of information resource transmission rate of communication network is

$$R_n = W \log_2 \left( 1 + \frac{G_n P_n}{\Gamma W N_0} \right) \quad (6)$$

Wherein,  $W$  represents the channel bandwidth of the communication network;  $G_n$  represents the channel gain in the time slot;  $P_n$  represents the transmission power of the wireless communication network;  $\Gamma$  represents the auxiliary constant;  $N_0$  represents the noise power density.

Then the energy required for information resource scheduling of communication network in time slot is

$$E_n^T = P_n \times \Delta T \quad (7)$$

Where  $\Delta T$  represents the length of each time slot.

Since there are many task types of cloud computing resources and applied resources, only tasks with matching types will be assigned to corresponding computing resources for task processing. In addition, multiple computing resources can work at the same time. The steps are as follows:

Step 1: after obtaining the energy data required for scheduling, place  $k$  ants on the energy samples required for scheduling, record the number of samples as  $M$ , and open the sample scheduling feature association pair at the same time;

Step 2: ants select the energy characteristics required for scheduling with transfer probability. At the same time, taboo table  $tabu_k$  and temporary pool  $temppool(k)$  are set,  $tabu_k$  is used to record the sample scheduling feature association pair selected by ant  $k$ , and application  $temppool(k)$  is used to save the probability value selected by ant [14]. Among them, under the limitation of energy required for information resource scheduling of communication network in time slot, the calculation formula of transfer probability is

$$p_{ij}^k(t) = \begin{cases} \frac{\alpha \cdot \tau_{ij} + \beta \cdot \eta_{ij}}{\sum_{r \neq tabu_k} \alpha \cdot \tau_{ij} + \beta \cdot \eta_{ij}} & j \neq tabu_k \\ 0 & \text{others} \end{cases} \quad (8)$$

Wherein,  $\alpha$  represents the weight parameter of pheromone quantity  $\tau_{ij}$ ;  $\beta$  represents the weight parameter of visibility parameter  $\eta_{ij}$ , which is calculated by  $1/d_{ij}$ , where  $d_{ij}$  represents the distance of a sample scheduling

feature association pair  $(i, j)$  among a plurality of pedestrians.

Step 3: in the process of path optimization, the population ant will leave the pheromone quantity  $\tau_{ij}$  in real time. The pheromone is the pheromone related to the scheduling order of the scheduled task. The scheduling order is the order in which the tasks in the taboo table are arranged to be processed by the corresponding computing resources. The pheromone here indicates whether a computing task may be assigned a specific order, In order to prevent the pheromone quantity of a path (Association pair  $(i, j)$ ) from increasing rapidly, resulting in the algorithm falling into local optimization, the initial pheromone quantity is set to  $\tau_{max}$  and the value range is  $\tau_{ij} \in [\tau_{min}, \tau_{max}]$ ;

Step 4: in the process of path selection [15-16], update the pheromone quantity based on local update rules. After the path selection, the pheromone update rule is changed to the global update rule. The global update rule expression is

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \Delta\tau_{ij} + \Delta\tau_{ij}^* \quad (9)$$

Wherein,  $\rho$  represents the pheromone quantity update coefficient;  $(1-\rho)$  represents the pheromone residue coefficient;  $\Delta\tau_{ij}$  means that  $(i, j)$  path pheromones are updated according to the path length ranking information of  $\sigma-1$  excellent ants, and the calculation formula is

$$\Delta\tau_{ij} = \sum_{\mu=1}^{\sigma-1} \Delta\tau_{ij}^{\mu} \quad (10)$$

Wherein,  $\mu$  represents the ranking of ants participating in path search in ant colony;  $\Delta\tau_{ij}^{\mu}$  indicates the amount of pheromone retained by the first ant on the associated pair.

Step 5: after determining the incidence matrix, record all the sample scheduling feature information to obtain the correct multi scheduling energy target feature data association results.

### C. Cloud Computing Virtualization Resource Scheduling based on Improved Ant Colony Algorithm

The original ant colony algorithm also has many shortcomings that are difficult to overcome by itself:

1) The ant colony algorithm framework operates on the basis of roulette algorithm. In the initial stage of solution space exploration, the utilization of prior knowledge is insufficient, resulting in a long time to search for a satisfactory solution. The certainty of the algorithm should be appropriately increased in the whole algorithm cycle [17-18].

2) After several rounds of search iteration, the evolution rate of the solution often decreases sharply, and even the basic gap between the found solutions is small, that is, the search stalls. The weight relationship between randomness and certainty should be dynamically adjusted according to different operation stages of the algorithm [19-20].

Therefore, simulated annealing algorithm is used to

optimize ant colony algorithm to overcome the above problems.

Combinatorial optimization problem is to solve the optimal solution. In the time period, set a non-negative objective function representing the solution goal on the ant colony path. The solid annealing process and combinatorial optimization problem have certain similarities, and the main advantages are as follows:

1) *Perform complex area search*: The algorithm is more suitable for searching in complex regions and obtaining regions with higher region values.

2) *Good parallelism*: The good parallelism of ant colony algorithm and simulated annealing algorithm can effectively solve various nonlinear problems. The good parallelism of the ant colony algorithm and the simulated annealing algorithm can solve various nonlinear problems effectively. The ant colony algorithm has very good time efficiency, but the ability to find the optimal solution needs to be improved, while the simulated annealing algorithm has very good ability to find the optimal solution, but the time complexity is high. By parallelizing the two, the advantages of both can be combined and the disadvantages of both can be compensated, and the final parallel algorithm not only has good ability to find the optimal solution, but also its time efficiency is high.

3) *Search with object functions*: Directly convert the objective function into fitness value to determine the search scope and direction in the next step.

Due to the slow development of physical state to low-energy state, in order to obtain better results, simulated annealing algorithm mainly selects the state with important contribution rate in the process of sampling.

Set the position representation state of the particle solid as the initial state, the energy is fixed, randomly form a small displacement change and produce a new state, set this state as an important state, the next state of the solid is sum, and the factor ratio corresponding to the probability ratio of the two states can be expressed as:

$$p(t) = \begin{cases} (E_i - E_j) \\ \frac{1}{Z(t)} \exp\left(-\frac{E_j - E_i}{k_B(t)}\right) \end{cases} \quad (11)$$

Where  $Z(T)$  represents the modified pheromone value on the ant path  $(i, j)$  at time  $t$ . It is set that the random number is a value formed by the random number generator, and the value range is within the range of  $[0,1]$ . Since  $p$  represents a value less than 1, assume  $p \geq A$ , indicating that it is sufficient to maintain the current state.

The resource types after optimized scheduling are divided into IO type, intermediate type and CPU type. When optimizing the scheduling, the resources are divided in combination with the running time sequence. The resulting resource division scheme is shown in Fig. 5.

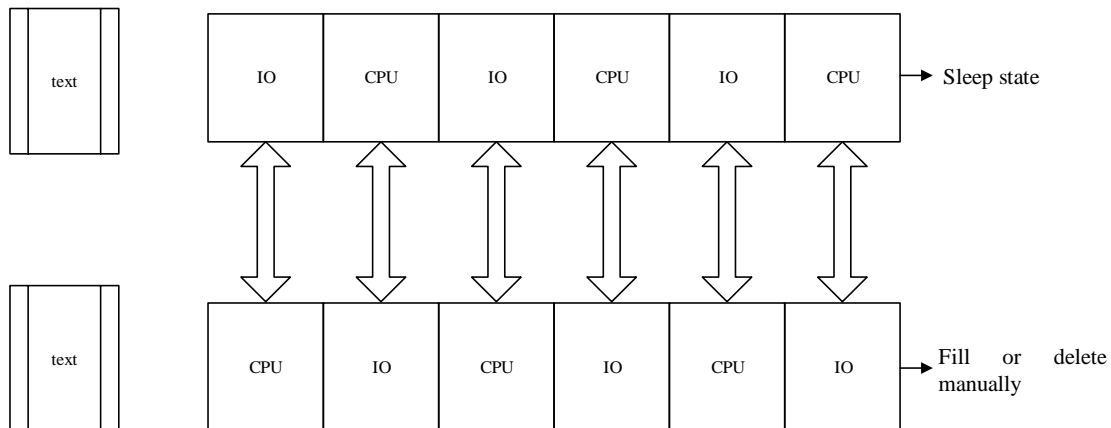


Fig. 5. Resource division scenario

As shown in Fig. 5, resource A and resource B consume the same disk (IO) and CPU, so both can be divided into intermediate types of resources. However, it can be seen from Fig. 4 that the time occupied by disk (IO) and CPU is completely crossed. Therefore, in the process of resource scheduling, the existence of AA or BB should be avoided as far as possible. In addition, missing and redundant data will be automatically filled or manually deleted in the process of resource optimal scheduling.

It is worth noting that the goal of the algorithm is to limit the number of physical hosts providing services in the cloud platform, and the trigger of any migration scheme is to migrate the virtual machine to an active physical machine.

However, when the current active physical machine cannot meet the requirements of the virtual machine to be migrated. The dormant physical machine will be activated as the destination physical host in the migration plan. A physical host is switched to sleep if and only if the virtual machines on it have been migrated. That is, when a physical host does not provide resources for any virtual machine, it will be converted to sleep state.

The important state is mainly set as the state with small probability acceptance and large current energy difference; In the high temperature state, the new state with large energy difference from the current state is accepted with a high probability and set as the important state.

In the actual application scenario of scheduling algorithm, due to the complexity and heterogeneity of cloud platform resources, the performance of physical hosts varies greatly, and the state of virtual machines relying on them will change dynamically with the deepening of task execution. In this case, it is easy to cause a large number of tasks to flow into the waiting queue of high-performance virtual machines and cannot be executed, resulting in service degradation and user experience degradation. On the other hand, there are a large number of idle virtual machines in the cloud platform, which are not favored by tasks due to limited performance, resulting in a great waste of cloud platform resources. As a result, the service life of overloaded machines is shortened, and the completion time of single tasks is not satisfactory. In order to solve this problem, it is necessary to analyze the needs of different users, calculate the matching degree of each demand corresponding to the task to the resource node, and add the matching factor  $Tm_i$ . Then the scheduling requirements in important states are expressed as follows:

$$Match_{ij} = \frac{1}{\sqrt{\sum_{i=1}^5 (Tm_i - Vm_j)^2}} \quad (12)$$

Set the mapping sequence  $Tm_i$  representing tasks and resources,  $Vm_j$  represents a sequence node. And the load balance degree can be expressed as:

$$Load(X) = \sqrt{\sum_{i=1}^n \left(1 - C_1 \frac{Y_i}{Lv_i}\right)^2} \quad (13)$$

Calculate the computing power of each resource node by weighting, that is:

$$Lv_i = w_1 \times n(v_i) + w_2 \times p(v_i) + w_3 \times r(v_i) \quad (14)$$

When pheromones accumulate to a certain extent, positive feedback suppression is needed, because with the deepening of iterative search, the total amount of pheromones rises to a certain threshold, which is prone to excessive positive feedback, that is, the influence of pheromones on ant colony covers the prior effect of heuristic function. In addition to ensuring the update speed of pheromones and highlighting the differences between high-quality paths and general paths, it is very important to comprehensively improve the load balance of resources. Therefore, the time span and load balance are set as the measurement indicators:

$$\begin{cases} F(X) = Load(X) \cdot Makspan(X) \\ Makspan(X) = \sum_{i=1}^n \sum_{j=1}^m s_{ij} \times E_{ij} \end{cases} \quad (15)$$

Above, mainly through simulated annealing algorithm for data center resource scheduling, the specific operation process is as follows:

- 1) Initialize all pheromones;
- 2) Setting up the main objective function and initializing each parameter;
- 3) When the ant starts the cycle, it needs to place all the tasks on a single resource.
- 4) Calculate the matching factor, set it as heuristic information, and select nodes in combination with the transfer probability.
- 5) Calculating the load balance degree and the value of the objective function, and the optimal solution of the ant colony algorithm.
- 6) The simulated annealing algorithm is used to optimize the path and update the pheromone.
- 7) Loop through the above steps until the maximum number of iterations is reached, then stop the operation and vice versa return step (3).

### III. SIMULATION EXPERIMENT

#### A. Experimental Data and Process

In order to verify the scheduling performance of the proposed method, comparative experiments are carried out. Provide 600 to 1000 cloud tasks in the resource submission window, and use the random number generator to randomly generate 50 waiting tasks with different requirements for cloud computing resources in the scheduling time window. It is estimated that the starting processing time is 0. Assuming that the time window is 40 minutes in length, the iteration is terminated under the condition that 6,000 ants all complete the iteration and the number of intelligent population is 100. When there is no change in the results of 10 consecutive iterations, the operation of the algorithm is stopped and an error is reported. Each set of data was analyzed 10 times and the average value was taken, and a parameter change was made to observe its impact on the delay time and solving efficiency of the task.

Hybrid cloud is a more advantageous infrastructure, which flexibly combines the internal capabilities of the system with external service resources, and ensures low cost. Hybrid cloud is a model that organically integrates public cloud and private cloud, allowing users to choose between the privacy of private cloud and the flexibility and cheapness of public cloud. Hybrid cloud can provide enterprises with more flexible cloud computing solutions. In the next few years, hybrid cloud will become the leading enterprise IT architecture, and more and more enterprises will benefit from hybrid cloud services. Considering the transport efficiency elements of cloud platform virtualization resources, a more commonly used cloud computing architecture is chosen as the experimental basis, as shown in Fig. 6.

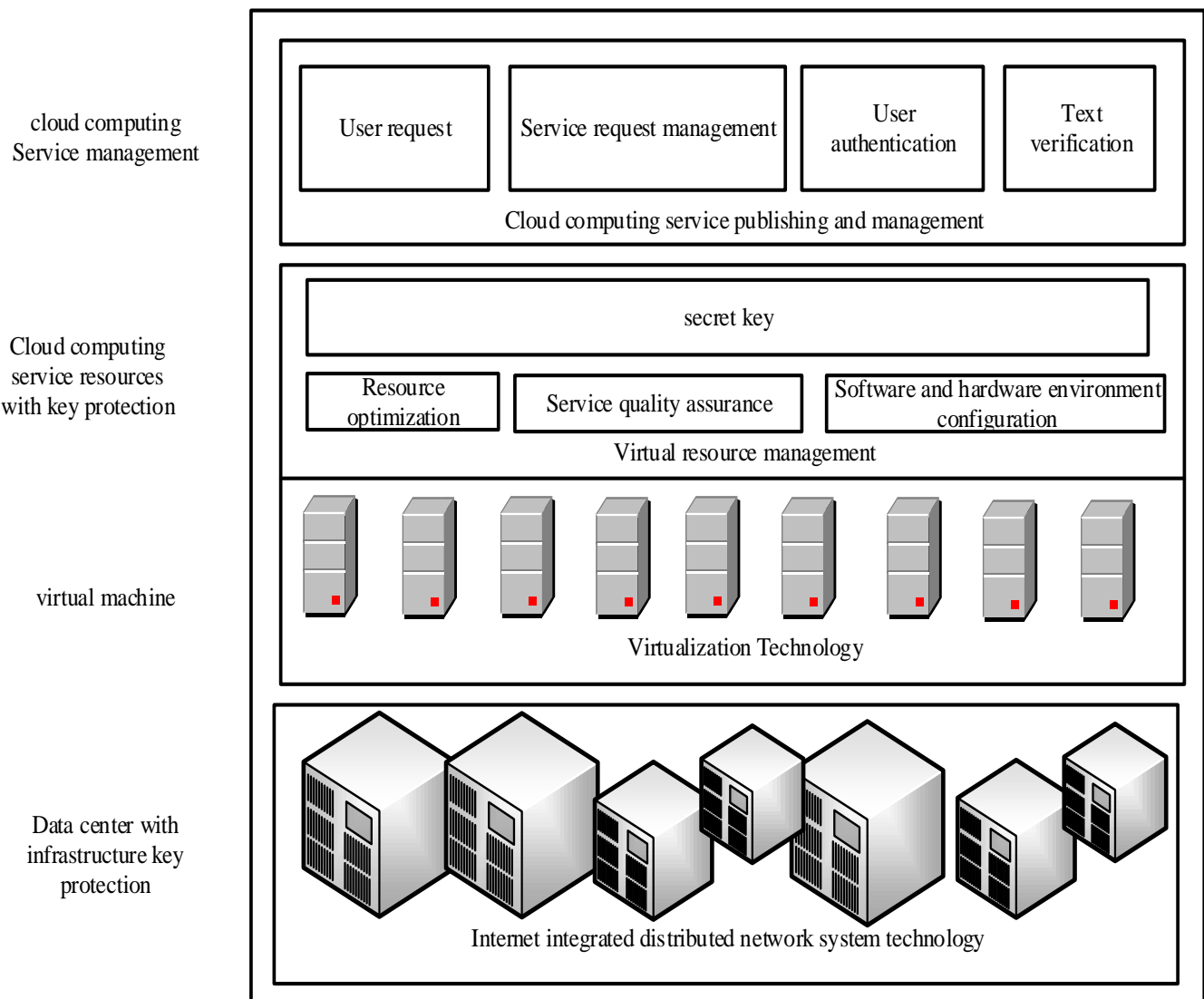


Fig. 6. Cloud computing architecture

In Fig. 6, according to the characteristics of multiplexed information, the structure of integrated distributed network and virtual machine is redeployed to ensure that the structure can completely load cloud computing virtualization resources.

The experimental parameters are shown in Table I.

TABLE I. EXPERIMENTAL PARAMETERS

parameter	analysis	numerical value
H	Total number of system storage nodes	6000
Size of File	Virtualization resource size	10GB
Backup Num	Number of backup copies	6000
Availability	System availability	>95%
P	Node availability	>95%

The simulation experiment goes through four processes, which are initialization stage, virtual resource access stage, virtual resource backup stage and virtual resource recovery stage. When the system is in the initialization phase, 6000 online nodes are built and node IDs are assigned. When the system is in the virtual resource access phase, the online phase needs to be divided into three categories: one is the service scheduling node, the other is the virtual resource backup node, and the third is the backup server. These three kinds of three wire nodes can achieve virtualized resource storage and backup through coordinated work. When the system is in the virtualization resource recovery stage, judge the node availability, and then use the backup server to recover the virtualization resources.

Several storage backup nodes are simulated in the MATLAB platform, the CPU speed and network bandwidth of the data center are configured, and the number and frequency of user information requests, as well as the start, end and peak time are set.

### B. Experimental Results

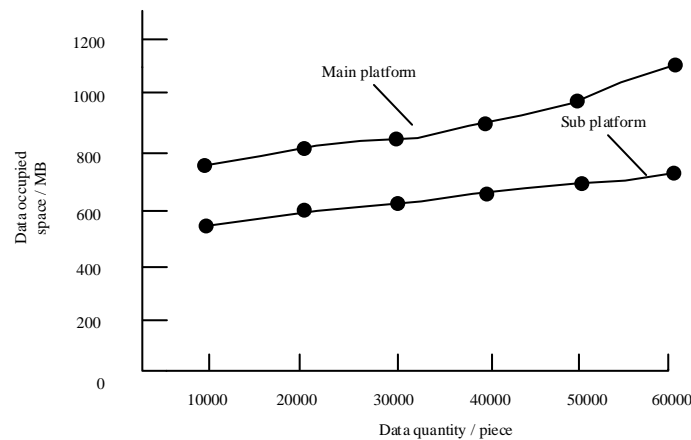
By comparing several different methods, taking the occupied space of the virtual resource scheduling process as the experimental index, and taking the methods of literature [2] and literature [3] as the comparison method, the simplicity of the application process of this method is verified. The experimental results are shown in Fig. 7.

As can be seen from the comparison effect of data occupation space shown in Fig. 4, the main and sub platforms have different storage capacity under the action of different data volume, but they are increasing with the increase of data volume. However, the initial occupation space of the method in this paper is the smallest and the growth fluctuation is low. This is because this method updates the pheromone quantity based on local update rules, that is, the pheromone quantity update rules are changed into global update rules to ensure that the task of type matching is accurately allocated to the

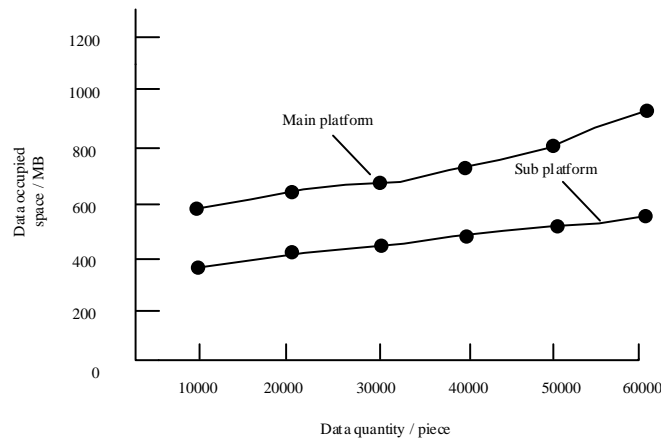
corresponding computing resources, which further increases the possibility of exploring a larger solution space.

Under different conditions, taking the load fairness test of cloud computing nodes as the experimental index, the comparison results of different methods are shown in Fig. 7.

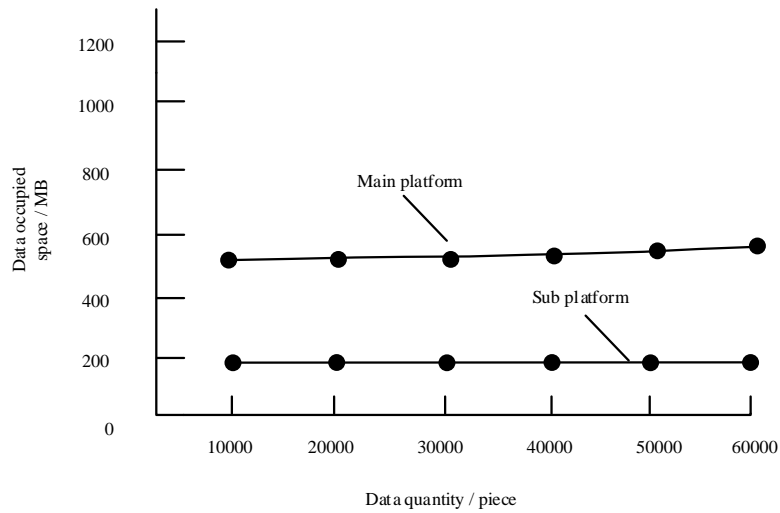
As shown in Fig. 8, the higher the load fairness coefficient, the better the storage effect. When the amount of backup data increases, the load fairness coefficients of other methods continue to decline, while the decline rate of this method is the lowest. This is because this method analyzes the needs of different users, calculates the matching degree of tasks corresponding to each demand to resource nodes, directly converts the target function into fitness value, and sets the time span and load balance as measurement indicators, Highlight the difference between high-quality path and general path.



(a) Literature [2] methods

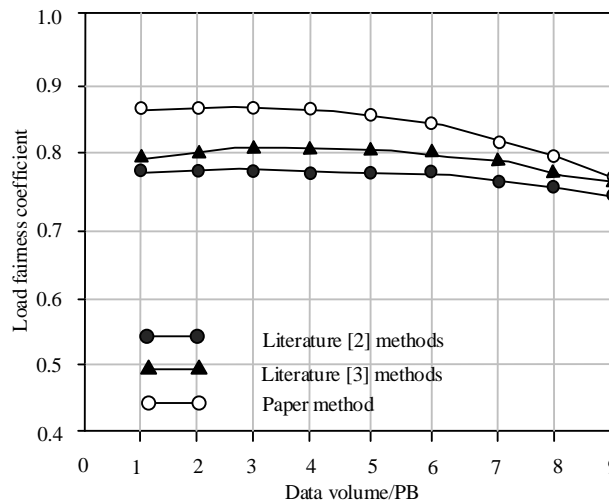


(b) Literature [3] methods

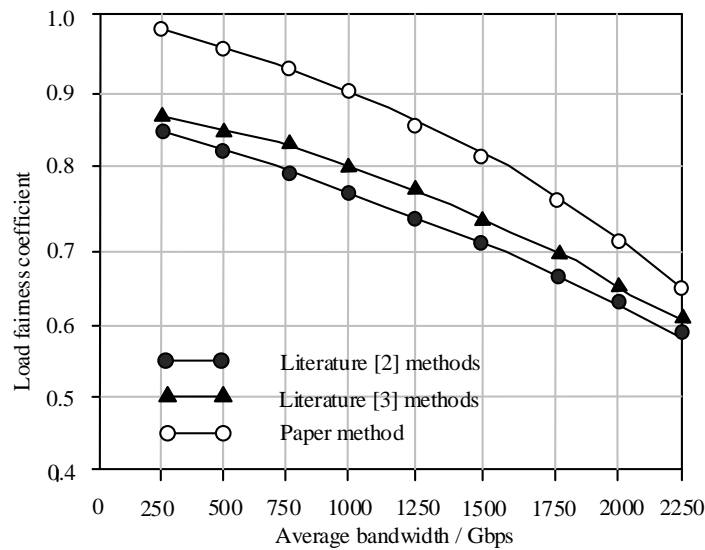


(c) Paper method

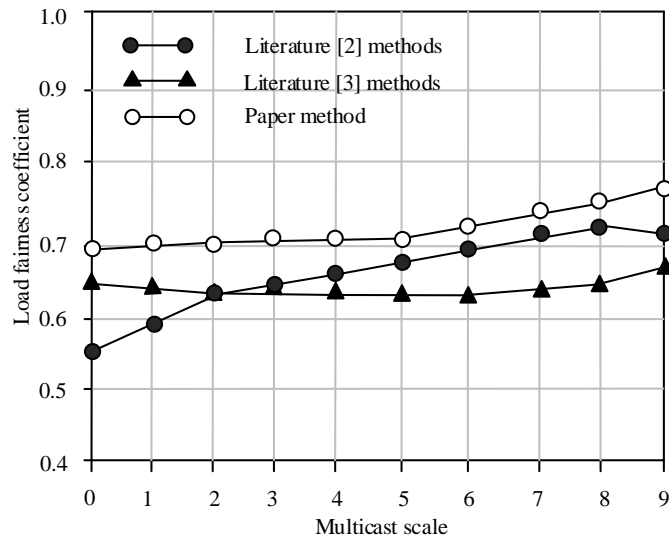
Fig. 7. Space occupied by different methods under main and sub platforms



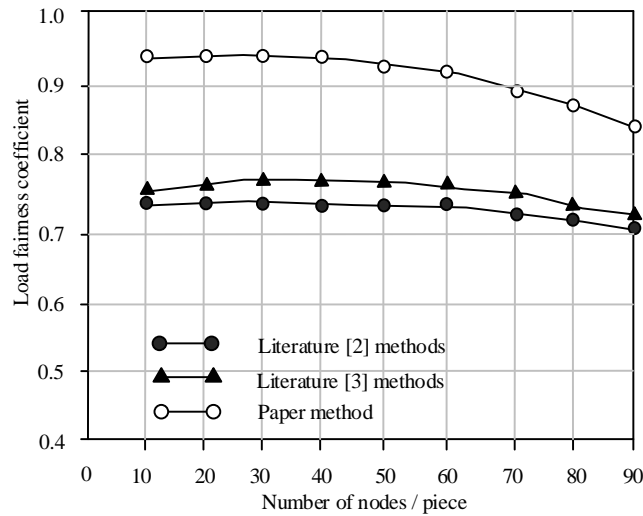
(a) More prior knowledge



(b) Complex area search



(c) Search stagnation



(d) Increase in the number of nodes

Fig. 8. Load fairness test results under different conditions

Calculate the physical machine node allocation rationality, resource balance, longest queue length and energy consumption of the three allocation schemes. The results are shown in Table II.

As can be seen from Table II, this method performs well in terms of Allocation Rationality, resource balance, longest queue length and energy consumption value. This is because

while constructing the communication network model of cloud computing platform, this method calculates the ratio of output power and input power of uplink resources at different times, and timely adjusts the scheduling demand according to the scheduling energy to reduce the scheduling energy consumption, effectively solve the problem of uneven spatial distribution of information of each data node in the cloud platform.



TABLE II. ALLOCATION RATIONALITY, RESOURCE BALANCE, LONGEST QUEUE LENGTH AND ENERGY CONSUMPTION

Solution method	Number of container / physical machine nodes	Rationality of distribution	Resource balance	Maximum queue length	Energy consumption value / W
Paper method	50/50	0.855	10.555	1	1000
	100/30	0.885	10.510	3	1500
	500/80	0.850	8.555	8	5000
	500/100	0.805	8.510	10	5000
Literature methods [2]	50/50	0.751	8.155	3	1500
	100/30	0.755	8.510	7	5100
	500/80	0.555	8.551	13	5500
	500/100	0.550	7.850	18	7000
Literature methods [3]	50/50	0.733	8.051	5	1500
	100/30	0.710	8.155	8	5550
	500/80	0.785	8.555	15	5000
	500/100	0.751	8.055	53	7500

#### IV. CONCLUSIONS

In order to shorten the space occupation of cloud computing virtual resource scheduling and improve load fairness, the study combines the advantages of simulated annealing algorithm with ant colony algorithm, and applies the hybrid algorithm to cloud computing virtual resource scheduling. The results of the empirical analysis show that the virtual resource scheduling process of the proposed method in the primary and secondary platforms occupies 535 MB and 196 MB of space, respectively; the load fairness coefficient of the method is around 0.85, both of which are better than the comparison algorithms. Besides, it is also found that the resource scheduling algorithm proposed in the study has more reasonable allocation, more balanced resources, shorter longest queue length and less energy consumption compared with the comparison algorithm. The above results illustrate that the resource scheduling algorithm proposed in the study has better scheduling performance, and the method can be used to better schedule the resources in the cloud computing platform and improve the resource scheduling efficiency, thus improving the computational efficiency and reducing the running time of the platform algorithm, thus making better and faster use of virtual resources and saving time. In the future research work, we should further strengthen the research on the resource scheduling efficiency and improve the performance of the cloud computing virtual resource scheduling method comprehensively.

#### V. DISCUSSION

The resource scheduling method proposed in the study combines the advantages of simulated annealing algorithm and ant colony algorithm, and compensates the disadvantages of both of them to each other, thus improving the overall algorithm performance. The parallel algorithm based on simulated annealing algorithm and ant colony algorithm has better scheduling performance in the process of resource scheduling in the cloud platform. In addition, the study also uses adaptive coding and modulation to extract the association

features of virtual resources. The study transforms the pheromone update rules into global update rules through the good parallelism of ant colony algorithm and simulated annealing algorithm to schedule the cloud computing virtual resources one by one in a specific order. It expands the scale of the cloud computing center cluster, improves the scheduling capability of the whole platform, makes full use of the cloud platform resources, and gives full play to the advantages of cloud computing. In the subsequent research, the algorithms that can make up for each other's defects can be combined to obtain a better hybrid algorithm, and the optimized hybrid algorithm can be applied to the scheduling of virtual resources using the optimized hybrid algorithm in order to achieve better scheduling performance of cloud computing virtual resources.

#### REFERENCES

- [1] W. Bo, J. L. Dang, Z. T. Li, H. F. Gong, F. Zhang, and S. Oh, "Modeling Analysis and Cost-Performance Ratio Optimization of Virtual Machine Scheduling in Cloud Computing," *IEEE Transactions on Parallel and Distributed Systems: A Publication of the IEEE Computer Society*, vol. 31, pp. 1518-1532, 2020.
- [2] A. Ramegowda, J. Agarkhed and S. R. Patil, "Adaptive task scheduling method in multi-tenant cloud computing," *International Journal of Information Technology*, vol. 12, pp. 1093-1102, 2020.
- [3] L. Abualigah, and A. Diabat. "A novel hybrid antlion optimization algorithm for multi-objective task scheduling problems in cloud computing environments," *Cluster computing*, vol. 24, pp. 205-223, 2021.
- [4] J. A. J. Sujana, T. Revathi, T. S. S. Priya, and K. Muneeswaran, "Smart PSO-based secured scheduling approaches for scientific workflows in cloud computing," *Soft computing: A fusion of foundations, methodologies and applications*, vol. 23, pp. 1745-1765, 2019.
- [5] M. Bansal, and S. K. Malik, "A multi-faceted optimization scheduling framework based on the particle swarm optimization algorithm in cloud computing," *Sustainable computing: Informatics and systems*, vol. 28, pp. 100429.1-100429.8, 2020.
- [6] V. Priya, C. S. Kumar, and R. Kannan, "Resource scheduling algorithm with load balancing for cloud service provisioning," *Applied Soft Computing*, vol. 7, pp. 416-424, 2019.
- [7] R. Valarmathi, and T. Sheela T, "Application of Particle Swarm Optimisation-R Factor Optimisation Algorithm for Efficient Resource Allocation and Scheduling in Cloud Environment," *Journal of*

- computational and theoretical nanoscience, vol. 17, pp. 1990-1998, 2020.
- [8] B. B. Naik, D. Singh, and A. B. Samaddar, "FHCS: Hybridised optimisation for virtual machine migration and task scheduling in cloud data center," IET communications, vol. 14, pp. 1942-1948, 2020.
- [9] S. Mostafavi, and V. Hakami, "A Stochastic Approximation Approach for Foresighted Task Scheduling in Cloud Computing," Wireless personal communications: An International Journal, vol. 114, pp. 901-925, 2020.
- [10] A. A. H. Al-Mahruqi, G. Morison, B. G. Stewart, and V. Athinarayanan, "Hybrid Heuristic Algorithm for Better Energy Optimization and Resource Utilization in Cloud Computing," Wireless Personal Communications, vol. 118, pp. 43-73, 2021.
- [11] M. Nanjappan, G. Natesan, and P. Krishnadoss, "An Adaptive Neuro-Fuzzy Inference System and Black Widow Optimization Approach for Optimal Resource Utilization and Task Scheduling in a Cloud Environment," Wireless Personal Communications, vol. 121, pp. 1891-1916, 2021.
- [12] H. O. Salami, A. Bala, S. M. Sait, and I. Ismail, "An energy-efficient cuckoo search algorithm for virtual machine placement in cloud computing data centers," The Journal of Supercomputing, vol. 77, pp. 13330-13357, 2021.
- [13] P. Partheeban, and V. Kavitha, "Versatile provisioning and workflow scheduling in WaaS under cost and deadline constraints for cloud computing," Transactions on Emerging Telecommunications Technologies, vol. 30, 2019.
- [14] R. G. Narendrababu, and K. S. Phani, "Regressive Whale Optimization for Workflow Scheduling in Cloud Computing," International Journal of Computational Intelligence and Applications, vol. 18, pp. 7-18, 2019.
- [15] L. Abualigah, and A. Diabat, "A novel hybrid antlion optimization algorithm for multi-objective task scheduling problems in cloud computing environments. Cluster Computing, vol. 24, pp. 205-223, 2020
- [16] X. Li, and H. Z. Sun, "Virtual Machine Matching Scheduling Simulation under Uncertain Network Load," Computer Simulation, vol. 37, pp. 363-366, 2020.
- [17] He W, Meng S, Wang J, Wang L, Pan R and Gao W. Weaving scheduling based on an improved ant colony algorithm. Textile Research Journal, vol. 91, pp. 543-554, 2021.
- [18] Zhao H, Zhang C and Zhang B. A Decomposition-Based Many-Objective Ant Colony Optimization Algorithm with Adaptive Reference Points. Information Sciences, vol. 540, pp. 435-448, 2020.
- [19] Li S, You X and Liu S. Multiple ant colony optimization using both novel LSTM network and adaptive Tanimoto communication strategy. Applied Intelligence, vol. 9, pp. 1-21, 2021.
- [20] Rueda R, Ruiz L, MP Cuéllar and Pegalajar M C. An Ant Colony Optimization approach for symbolic regression using Straight Line Programs. Application to energy consumption modelling. International Journal of Approximate Reasoning, vol. 121, pp. 23-28, 2020.

# Deep Analysis of Risks and Recent Trends Towards Network Intrusion Detection System

D. Shankar<sup>1\*</sup>, G. Victo Sudha George<sup>2</sup>, Janardhana Naidu J N S S<sup>3</sup>, P Shyamala Madhuri<sup>4</sup>

Research Scholar, Department of Information Technology,

Dr. M.G.R. Educational and Research Institute Chennai, Tamil Nadu 600095, India<sup>1\*</sup>

Professor, Department of Computer Science and Engineering,

Dr. M.G.R. Educational and Research Institute Chennai, Tamil Nadu 600095, India<sup>2</sup>

Assistant Professor, Department of Computer Science and Engineering,

Vishnu Institute of Technology Bhimavaram, Andhra Pradesh, 534202 India<sup>3,4</sup>

Assistant Professor, Department of Computer Science and Engineering,

Vishnu Institute of Technology Bhimavaram, Andhra Pradesh, 534202, India<sup>4</sup>

**Abstract**—In the modern world, information security and communications concerns are growing due to increasing attacks and abnormalities. The presence of attacks and intrusion in the network may affect various fields such as social welfare, economic issues and data storage. Thus intrusion detection (ID) is a broad research area, and various methods have emerged over the years. Hence, detecting and classifying new attacks from several attacks are complicated tasks in the network. This review categorizes the security threats and challenges in the network by accessing present ID techniques. The major objective of this study is to review conventional tools and datasets for implementing network intrusion detection systems (NIDS) with open source malware scanning software. Furthermore, it examines and compares state-of-art NIDS approaches in regard to construction, deployment, detection, attack and validation parameters. This review deals with machine learning (ML) based and deep learning (DL) based NIDS techniques and then deliberates future research on unknown and known attacks.

**Keywords**—Network; dataset; communication; intrusion detection system; attacks; deep learning; machine learning

## I. INTRODUCTION

The Internet is ubiquitous everywhere that plays a vital role in daily activities. It almost pervaded every field of business activities, from retail to space research for storing and sharing confidential data [1]. A vital concern in businesses is to secure the network from external disturbances. The advancement of new communication technologies and services increases the number of interconnected devices [2, 3]. The escalation of Internet based systems also increases vulnerabilities and virtual attacks on the network. Cyber-crimes on the Internet are rising exponentially with the rapid growth of technology. The cyber-attacks are severe attacks on the network, and a failure in security may cause misleading information. Thus the network cannot meet its goals with the improper IDS [4]. Moreover, communication in the network is affected by individual hackers and malicious behavior, which seriously impacts the system [5]. Different types of attacks occur in the network, such as wormhole attacks, denial of services (DOS), black hole attacks, and flooding attacks. All these attacks prevent the user from accessing network

resources and services by dropping or adding unnecessary packets [6]. Additionally, variations in malicious software will increase the threats to the network through security breaches.

An intrusion detection system (IDS) has drawn considerable interest among academics since it protects the network against inside and outside attacks, a proactive ID tool. The IDS is a security application used after the antivirus software and firewall. The IDS alters the network when any faults occur in the system. The IDS not only detects intrusions but also stores the network traffic in the system [7]. The IDS alerts the system when the intrusion occurs at the host and network levels [8]. The available IDS have many limitations having difficulties coping with the cyber-attacks due to their dynamic nature. The IDS cannot be easily adaptable; thus, it cannot detect new malicious attacks [9].

Furthermore, the massive network size and the larger number of applications handled by the node have resulted in the IDS becoming a challenging task owing to the generation of huge amounts of data. The massive data generated in the network is communicated with other nodes inside the network will create new attacks [10]. Hence, several methodologies have been developed in the past few issues to deal with these issues.

Deep learning (DL) and machine learning (ML) approaches have recently been widespread in several applications. The DL based methods have functioned with the neural networks containing numerous layers. In ML, the algorithms analyze the data to solve the issues by learning and pattern discovery. These methods are found to be excellent in higher detection accuracy and better performance on dataset testing; hence, these methods are included in IDS. Thus this paper has reviewed traditional methods to ML and DL based methodologies for ID by considering different types of faults [11]. Recently, the DL and ML based approaches gaining more attention in the IDS context. In fact, the DL is used for feature extraction, selection and classification. The DL networks carry the learning process in supervised and unsupervised manners. The Internet is omnipresent in all fields, which is inevitable in network communication. Thus the applicability of IDS in networks is considered a vital

research topic for enhancing network security against attacks. However, several authors have made a review this topic. Some of the recent papers of review of IDS are listed below:

- A review on DL based IDS is conducted in [12, 13] that only examines the DL based methods in IDS, but that paper did not focus much on network attacks.
- Recently, the IDS in the Internet of Things (IoT) have been reviewed, and some solutions are suggested to overcome the issues [14].

The existing review papers provide a deep insight on DL, ML or existing approaches individually. However, they do not analyze up-to-date methods. Moreover, different types of datasets are not focused in deep. Apart from these recent reviews, this paper is intended to focus on the up-to-date methods for IDS thus it reviews traditional methods. The major aim of this paper is to review various IDS methods, ID behavior, performance efficacy and problems associated with IDS methods. Furthermore, the leading methods in the IDS are also discussed with their advantages. Although several techniques have been developed in this area, no single method covers all types of network attacks. In the paper, innovative content is added in terms of up-to-date methods by including all traditionally available methods for IDS in the network by considering different types of attacks in the network. Thus, this review has provided the growth of IDS and the issues the developed methods face.

The contribution of this paper is summarized as follows:

- To study the network security issues and point out the necessity of an IDS as a solution for detecting attacks.
- To emphasize attacks in networks, then provide the key role of IDS for network security.
- Highlight the challenges and advantages of the current works in intrusion detection.
- To bring an effective solution to overcome the issues associated with recently developed methods.
- Provide an excellent approach for detecting the harmful unknown intrusions in the network by resolving the issues in existing and newly developed methods based on the analysis.

The organization of this review paper is given as follows; section-2 deals with several attacks in the network, and Section 3 elaborates on the detection methods and their pros and cons. Section 4 explains the limitation of NIDS. Section 5 includes the conclusion and future scope for network ID.

## II. ANALYSIS OF ATTACKS IN THE NETWORK

Data transmission through the network may be subject to various threats and safety risks. Hence, it is often unable to be secure with new techniques. It is very important to analyze the intrusion and attacks in the network to secure it from failure. Malicious nodes in the network create a routing attack. Different security threats and network attacks are discussed below.

### A. Security Threats in Network

The network should meet some security measures to prevent intrusions from outside the network. An unauthorized user can enter the network if the organization fails to meet sufficient security concerns like prediction, securing and isolation. On the other hand, data management in the network is affected by a lack of bandwidth management. Because data communication in the network depends on bandwidth resource utilization, thus the excess utilization of bandwidth in the network will lead to bandwidth wastage. The severe issue in network transmission is data confidentiality. The lack of an encryption mechanism will lead to data leakage in the network [15]. Network attacks such as snooping, traffic, modification, denial of services and injection of intrusions are common in the network [16].

### B. Active and Passive Attacks

Attacks in the network are classified into two types: active and passive attacks. The intruder intercepts the data in the network during a passive attack, whereas, in an active attack, the intruder initiates a command to disturb network operation.

- Spoofing: In a spoofing attack, the sender can change the network topology due to the miss-present data of a malicious node.
- Modification: In this kind of attack, the data route is modified by a malicious node so that the message may be sent through a long route. Data routing through long-distance leads to communication delays [17].
- Wormhole attack: In a wormhole attack, the data transmission between two nodes is affected due to malicious nodes. This attack fakes a route instead of the short route in the network; thus, distance gets confused [18].
- Sinkhole: The sink-hole attack attracts network traffic through false routing metrics. The sink-hole attack invites many attacks in the network, sending false information [19].

- Denial of service attack: Most communication networks have limited energy resources; hence, they do not cope with sophisticated safety technologies. One of the common attacks in network communication is a DOS attack. The DOS affects almost all network layers, thereby disabling the proper functioning of the network. The DOS also abuse data in the network by allowing multiple attackers. The DOS attacks may lead to poor network performance, spam messages, packet delay and loss [20].
- Sybil attack: This attack uses fake identities to transmit data on the network.
- Traffic analysis attack: this kind of attack examines node behavior, network traffic and length of the message.
- Eavesdropping: Eavesdropping on ongoing communication may cause information leakage on cryptography or connection [21].
- Monitoring attack: The intruder can read the confidential data but cannot modify the data.
- Remote to Local (R2L) attack: The attacker gets access to the network without an account in the R2L attack, considered a critical attack in the network. This type of attack depends on host level and network-level features.
- User to Root (U2R) attack: The U2R needed semantic data that was hard to capture at the initial stages. U2R commonly happens in content-based fields where the attacker starts as a normal user and then becomes a superuser to abuse the network.
- Probe: It is illegal to gather network information about several services and sources to violate network security.
- Man-in-the-middle (MITM) attack: In the MITM attack, the attacker can access both ends of the communication channel, thereby manipulating the data. In such attacks, the attacker tries to initialize secure communication by sending messages. Finally, the attacker gets access to encrypt all the messages in the communication channel [22].
- Malware attack: In this kind of attack, the evade signature is matched by dynamically modifying the code. Due to the different purposes of the attackers, it becomes difficult to detect the attack.
- Phishing attacks: Spam e-mails are created to advertise a product, which is slightly modified to harm the user is known as phishing attacks financially. Phishing crimes are created by making a fake website the same as the original website and then creating fraudulent offers to the user [23].
- Distributed denial of service: This attack disrupts the normal traffic of the targeted server or network by overwhelming the target.

### III. BACKGROUND OF INTRUSION DETECTION SYSTEM (IDS)

Intrusions in the system are noticed by continuous observation of the system, in which the system administrator examines intrusions through user activities. The main aim of IDS is to examine the intrusions in the system. However, it is a difficult task, in fact, the IDS has not examined intrusions at all; that only examines the symptoms of intrusions. The symptoms gathered by the IDS are referred to as manifestation, where the collected evidence is not sufficient for detection, then the system can't access the intrusions at all. As the network attacks increased with time, the early form of intrusion detection methodologies is not scaleable.

The IDS is a significant tool for the network to examine security breaches. The IDS continuously monitors the network traffic entering and leaving the system, thereby examining network intrusions. The communication in the network is carried out through wire and wireless mediums. Thus, the network attacks increased dramatically in the network. Moreover, technological advancements increase new attacks on the network. The intrusion of attacks common in new computing environments such as wireless sensor network, fog computing, e-healthcare and cloud computing. IDS is vital security component that enables the computer network for IT organization. Hence, it is needy to construct IDS to defend the network from attacks. To provide a sufficient solution for IDS, it is important to investigate the up-to-date methods.

#### A. Principles of IDSs

IDS is nothing but the process of observing the events that occur in the system and then examining the intrusions through these events [24]. Intrusions can take several forms routing attacks, sniffer attacks, U2R attacks, man-in-the-middle attacks, Dos/DDos, and cyber-attacks [25, 26]. One of the major tasks of IDSs is to separate harmful attacks from normal traffic in an efficient manner. Connecting to a wrong system by wrongly typing the address is an example of normal traffic, which is considered a threat. Due to the timely detection of intrusions, IDS protects the system or network from failure. The major functions offered by IDS are listed below.

- Provide timely alarm while detecting threats in the system.
- Take necessary action to respond to the alarm detected [27, 28].

Factors to be considered for effective IDSs are summarized below:

- Robustness of the system
- Speed of detection
- Maximum detection rate.
- The minimum detection rate of normal traffic is a threat.
- Reduction in the system requirement, which includes hardware and software
- Accuracy in detecting threat location

- Integration with other technologies.

1) *Data collection & recording*: Initially, the network data are gathered to create the profiles of normal data compared with the observed data of the network. After initializing the profile of normal data, the network data are collected to check the intrusions. Since the volume of collected data is huge, creating small groups as vectors is useful in detecting intrusions.

2) *Identification of harmful intrusions*: Harmful threats in the network are observed by analyzing the data. In this step, the host and network data were examined for accurate detection of intrusions. In this stage, the undesired actions outside the network are recorded to detect intrusions.

3) *Alert*: In this stage, the condition of the network is examined by the judgement based on evaluated data, which examines whether intrusion occurred or not. Once it is acknowledged that the threat has happened, the IDS immediately communicate it with the administrator. Few IDSs can control attacks by utilizing network resources. Communication of threat occurrence is performed using various platforms like e-mails and messages in the user interface [29, 30].

#### IV. DETECTION METHODOLOGIES

The ID methodologies are classified a:

- IDS types by detection technique
- IDS types by monitored platform

##### A. IDS Types by Detection Technique

Further, the detection technique is categorized as,

- Anomaly based model
- Specification based
- Hybrid methods

1) *Anomaly-based (anomaly detection)*: Anomaly-based IDS detection approaches compare observed activities with definitions to predict malfunctions. Normally, anomaly-based detection has some rules to define basic network functioning based on the intrusions observed. Sultan et al. [31] proposed an anomaly based method using a variational encoder for the ID in the network. The intrusion was detected through a semi-supervised learning approach and unsupervised deep learning (DL) methods. Variational autoencoder (VAE) and autoencoder (AE) were employed using flow features to detect unknown attacks. The area under receiver operating characteristics (ROC) was calculated using these methods and compared with a one-class (OC) support vector machine (SVM). AE and VAE were used, and OCSVM were trained through semi-supervised learning. To detect intrusion on flow based data, that approach uses DL methods. But that proposed method increases the false alarm rate.

Whereas, Shubhra et al. [32] proposed an adaptive scheme that combines both the adaptive grasshopper optimization

algorithm (AGOA) and the ensemble of feature selection (EFS) method for identifying attacks. The EFS ranked the attribute from a selected subset of attributes having higher ranks, and the AGOA was adopted to find significant attributes from datasets. SVM was used as a fitness function to increase classification performance and efficient feature selection; the proposed method decreased proficiency. Hence to improve the classification, Bayu et al. [33] proposed a Two-Stage Classifier based IDS (TSE-IDS) to detect attacks. In order to minimize the feature size of datasets, hybrid particle swarm optimization (PSO), ant colony algorithm (ACO), and genetic algorithm (GA) were used. In addition, a reduced error pruning tree (REPT) classifier is used for improving classification performance and feature selection.

In order to improve dynamic network performance, Nguyen et al. [34] proposed anomaly-based network IDs (NIDS) using DL to consider false-positive rates and the unavailability of labelled data. The proposed method used Restricted Boltzmann Machines (RBM) and Autoencoder. The staked auto encoder performed better than RBM but consumed more time due to higher computation. Future work is needed to improve speed and reduce the oscillations in the slope of training error. Roshan Kumar and Deepak Sharma [35] proposed a hybrid ID algorithm focusing on reduced time consumption. That proposed method was developed by combining anomaly and signature-based approaches. That proposed method was more effective for detecting more attacks. But that proposed method had the drawback of higher time consumption.

TABLE I. ANOMALY-BASED IDS

Author	Method	Datasets	Outcome	Advantages	Drawbacks
Sultan et al. [31]	Variational Autoencoder	NSL-KDD	AUC – 0.7596	The detection rate is higher	A false alarm is more
Shubhra et al. [32]	Adaptive scheme	ISCX 2012	A-99.13 DR-99.23 FPR-0.067	predict the networks traffic behavior accurately	Decrease in proficiency
BAYU et al. [33]	Two-Stage Classifier Ensemble	NSL-KDD and UNSW-NB15	A – 85.8% S – 86.8% P-91.60% DR-88%	Improved precision metric and accuracy	Multi-class classification on problem
Nguyen et al. [34]	NIDS	KDDCup99	T-240sec	Effective in detecting and classifying intrusion into five groups.	Higher oscillation in the slope of training error
Roshan et al. [35]	HyINT	KDDcup99	T-14.773	Ability to detect unknown attacks	Time-consuming process

Above mentioned methods for anomaly-based IDS methods for network ID and their outcome and advantages are illustrated in Table I. In every table, the performance measures are indicated by Accuracy (A), Detection rate (DR), Precision (P), F-score (F), Connection rate (CR), Learning rate (LR), True Positive rate (TP), False Positive rate (FPR) and Ability to avoid misclassification (AUC), false alarm rate (FAR), false positive (FP).

2) *Specification-based NIDS*: Specification-based IDS uses manual specifications to detect network intrusion. Various specification-based methods are listed below.

Anhtuan Le et al. [36] proposed Routing Protocol (RP) for Low Power and Lossy Networks (RPL). That proposed method was constructed by a semi-auto profiling technique that generates high-level abstracts. That proposed approach includes protocol states and transitions on statics executed on several IDS rules. The power consumption was minimized by eliminating the overhearing of communication. Furthermore, the RPL information object and Information Solicitation (DIS) were introduced to alleviate the synchronization issues. But that the proposed method needs to be improved to detect internal threats.

Hence, Herson Esquivel-Vargas et al. [37] proposed a specification-based IDS using the BACnet protocol to enhance the detection rate. In that approach, fully automated deployment of IDS through BACnet protocol was used. In that protocol, the certified devices were demanded to document, representing network behavior. The prototype of that protocol was executed passively, and the attacks were detected on a single BACnet packet. But that proposed method needs to be improved to increase network security.

Above mentioned methods for specification-based IDS, along with its outcome and advantages, are illustrated in Table II.

TABLE II. SPECIFICATION-BASED METHODS FOR IDS

Author	Approaches/modules	Application	Outcome	Advantages	Drawbacks
Anhtuan Le et al. [36]	semi-auto profiling technique	RPL-Based Network Topology	OH-6.3%	Higher energy efficiency	An extension of IDS is needed for detecting internal threats
Herson Esquivel-Vargas et al. [37]	synthetic traffic	BACnet Protocol	P-99.85% R-99.57%	Attacks can be detected during real and synthetic traffic	System security needs to be improved

3) *Hybrid techniques*: Hybrid techniques are developed by combining two methodologies for ID. Thus, the hybrid methods overcome the drawbacks of single approaches. Various hybrid approaches for NIDS are given below.

Further improving network performance and addressing the issues of RPL, Areej Althubaity et al. [38] proposed an architecture by combining a centralized model with a sink and

a distributed module with RPL. That hybrid method was called Authenticated Rank and Routing Metric (ARM). That proposed approach validates the legitimacy of data transferred through RPL control data while constructing the route. Sajad Einy et al. [39] proposed hybrid anomaly and signature-based NIDS for enhancing network security. Suricata IDS was adopted with a neural network model to examine the network's malicious behaviour in that proposed model. The signatures of different types of attacks were examined by the neural network (NN) model. Then, the output from NN was given to Suricata IDS. In addition, an open-source blacklist internet protocol was adopted in that system.

That proposed method may create additional attacks in the network. Hence to avoid additional errors in the network, MavraMehmood et al. [40] proposed a hybrid approach by combining SVM and an adaptive neuro-fuzzy interference system (ANFIS). That paper adopted min-max and data transformation methods for data pre-processing. Then, optimum features were chosen by the random forest recursive feature elimination technique. Afterwards, ANFIS and SVM were used for ID and classification, respectively.

Above mentioned hybrid methods for IDS, along with its outcome and advantages, are illustrated in Table III.

TABLE III. HYBRID METHODS FOR IDS

Author	Method	Approaches/modules	Outcome	Advantages	Drawbacks
Areej Althubaity et al. [38]	hybrid specification-based IDS	Centralized and distributed models	Power consumption is 1966.711 mW	Extra overhead is minimized	-
Sajad Einy et al. [39]	Hybrid anomaly and signature-based IDS	Suricata IDS	A-96.11%	Different attacks are detected	Lower accuracy
MavraMehmood et al. [40]	ANFIS and SVM	random forest recursive feature elimination technique	A-99.3% Sp-0.998% P-0.999% R-0.992% F-0.995%	Higher accuracy	Higher cost

### B. IDS Types by Monitored Platform (Data Source)

The monitor-based IDS methods are classified as:

- Network based IDS
- Host-based IDS

1) *Network based IDS (NIDS)*: A NIDS detects harmful traffic on a network requiring promiscuous network access to examine the traffic. The IDS used some components to protect the network from attacks. The NIDS management console and management server were secreted from the remaining network. Therefore, the attacker cannot determine the location of the components.

Basant Subba et al. [41] proposed an artificial neural network (ANN) for IDS. The ANN optimization techniques

minimize the computational overhead and maintain high-level performance simultaneously. ANN consists of more interrelated nodes collaborating near each other to make a solution an exact task. ANN depend IDS model was suitable for real-time consumption. That proposed method requires more processing time. Hence to reduce the time consumption, Norbert Adam et al. [42] proposed NN based IDS to identify cruel behavior in the network. Ethernet taps were used to divide the signals and then distribute one branch to the original objective and another branch to the IDS. That technique tested for the nMap scanning attack, UDP flood attack, SYN flood attack and non-hateful statement.

Vrushali D.Mane & SN Pawar [43] proposed a back propagation ANN (BPANN) algorithm for detecting various network attacks to reduce FAR. The main objective of those methods was safeguarding the complete data with the support of a supervised neural network. In that approach, the neural network only used significant features of the KDD 99 dataset. The system performance was analyzed using 10% of the data from the KDD 99 dataset. The KDD training dataset contains a large number of signal connection vectors. The neural network needs a lot of time to test and train all datasets.

Further, Hossein Gharaee & Hamid Hosseinvand [44] proposed a GA and SVM for feature selection to improve network performance. GA drops the data dimension similarly, improving true positive detection and reducing the FP detection. In the feature selection method, input was considered traffic data that produces features (chromosomes) and then chooses chromosomes with higher categorization precision. SVM reduced computational time for training and achieved a low FPR with high accuracy. In that technique, KDD CUP 99 and UNSW-NB15 datasets were used for the testing. That proposed model was not suitable for large datasets.

IDS was a valuable device for analyzing and detecting malicious activities in the cloud network, and that was employed in organizations, enterprises and is important in cyberspace security. Kai Zhang et al. [45] proposed an intrusion action based correlation framework (IACF) for analyzing and correlating malicious behaviors in networks. IACF was used for improving the procedure of action extraction, scenario discovery and alerting. IACF alerts the network depending on the conception of intrinsic tough correlations. The sequence pruning algorithm (SPA) decreases false-positive impact and constructs the correlation. The IACF technique was used for predicting intrusion behavior depending on correlation graphs.

Basant Subba et al. [46] proposed a game theory (GT)-based false alarm (GTBFA) for reducing the FPA in signature-based IDS. A high FPA rate led to considerable utilization of network assets for monitoring against useless network fear. That proposed method links IDS alarms with network behavior to reduce the FPA rate of IDS. GTBFA uses more malicious activity scanners to scan the cloud network and generate a threat summary of the network. DARPA and IITG lab network datasets were used to reduce the FPA rate of IDS. A GT procedure was used to develop the network's sensible vulnerability set (SVS). This proposed method offer reduces

the detection rate of critical vulnerabilities. At the same time, network classification had long-term problems such as irrelevant features and redundancy. Priyadarsi Nanda et al. [47] proposed the least square (LS) SVM for IDS for feature selection. The KDD Cup 99, NSL-KDD and Kyoto 2006+ datasets were used to examine the performance of the LSSVM based IDS. Flexible, mutual information feature selection (FMIFS) was an efficient feature selection algorithm to decrease similarity. SVM was able to solve the binary classification problems.

Mobile malware could direct to some cyber security threats such as installing backdoors, stealing sensitive information, sending premium SMSs and ransomware attacks. The antivirus systems were not able to detect the advanced threats. Therefore, there is an additional layer of safety on the network to protect the users from threats. Sanjay Kumar et al. [48] proposed a machine ML dependent on NIDS. ML classifiers were built using a dataset consisting of labeled instances of network traffic features created through some malicious. NIDS was capable of identifying malicious traffic effectively where antivirus created false negatives. ML classifiers were used as the most efficient and traditional antivirus in that technique. The ML model was integrated into traditional IDS for detecting advanced threats and decreasing false positives.

Communication systems had two major challenges: privacy of user-specific data and computer security. The increasing uses of Internet connected devices had a certain increased to more number of vulnerabilities, which integrated assault on devices. Ayyaz-UI-Haq Qureshi et al. [49] proposed a novel random NN depending on IDS (RNN-IDS) for detecting malicious activity. The RNN architecture consists of one way to pass the signals or information where data or signal transfers from the input layer to the hidden layer. In that technique, performance was estimated by training dissimilar numbers of input and hidden layer neurons through a learning tariff on standard NSL-KDD datasets for binary classification. The gradient descent algorithm (GDA) was used for classifying the binary class of NSL-KDD datasets. A large number of input neurons were needed to improve detection accuracy. Moreover, the proposed method detects only a few attacks.

Thus, Sanchit Nayyar et al. [50] proposed an LSTM based ML approach for examining intrusions in the network. That proposed approach was tested on the CICIDS2017 data set with several attacks such as Web based Brute Force, DoS Hulk, DoS sloworis, DoS GoldenEye DoS slowhttpstest, DDoS LOIT and Patator based attacks. The neurons in the input and output layers were 77 and 2; 12 hidden layers were presented. That proposed approach requires an efficient algorithm for better classification. Nuno Oliveira et al. [51] suggested a sequential approach based on NIDS.

Moreover, the performance of multilayer perception (MLP), random forest and long short term memory was estimated on the CIDDS-001 dataset. That proposed model was evaluated on single flow and multi-flow. Above mentioned methods and their advantages are illustrated in Table IV.



TABLE IV. NETWORK BASED IDS

Author	Method	Datasets	Classifier	Outcomes	Advantages	Disadvantages
Basant Subba in 2016 [41]	ANN	NSL-KDD	ANN	A-95.05% DR-95.05%	High accuracy and detection rate	More processing time for large neural network
Norbert Adam in 2017 [42]	NNIDS	Train.txt	ANN	CR-1 LR-0.5	Recognize learned malicious activities	Spanning port per switch was allowed
Vrushali D.Man in 2018 [43]	BPANN	KDD 99	ANN	A-98.0% DR-92.80%	Minimize FAR	Sensitive to noise data
Hossein Gharaee in 2016 [44]	GA and SVM	KDD CUP 99 and UNSW-NB15	SVM	A-99.05% TP-98.47% FP-0.4%	High accuracy and low FPR	SVM was not suitable for large datasets
Kai Zhang in 2019 [45]	IACF	LLDOS 1.0	SPA	A-90%	Efficient in alert correlation	Less information in the infiltration scenario
Basant Subba in 2016 [46]	GTBFA	DARPA and IITG Lab	SVM	A-98.55% DR-91.87%	High accuracy	The detection rate was low against non-critical vulnerabilities
Priyadarsi Nanda in 2016 [47]	LSSVM	KDD Cup 99, NSL-KDD and Kyoto 2006+	SVM	A-99.94% DR-98.93% FR-0.28%	Better accuracy and low computational cost	Not suitable for unbalanced sample distribution condition
Sanjay Kumar in 2016 [48]	ML	KDD99, DARPA 1998/1999 and ISCX 2012	ML classifier	A-99.4% DR-82% TP-99.6% FP-1.8%	Detect threats with high accuracy	Classifier modelled using few malware functions
Ayyaz-Ul-Haq Qureshi in 2018 [49]	RNN-IDS	NSL-KDD	GDA	A-94.50% P-98.9% D-95.3% TP-95.31% FP-1.28%	High precision value	More number of input and hidden neurons are required for better efficiency

Sanchit Nayyar et al. [50]	RNN-LSTM	CICIDS 2017 data set	RNN	A-96%	More attacks are founded	Need to develop an effective classification algorithm
Nuno Oliveira et al. [51]	Sequential approach	CIDDS-001 dataset	RF, MLP and LSTM	A-99.94% F-91.66%	Anomaly detection improved by sequential flow	Need to improve network optimization

2) *Host-based IDS (HIDS)*: Host-based IDS protects our systems from many harmful intrusions occurring in the network. The host-based approach detects normal behavior by sequencing the system call. System sequences in the network create the sequences.

Chawla et al. [52] proposed Convolutional Neural Network (CNN) - Gated Recurrent Unit (GRU) language model for the just released Australian Defense Force Academy Linux Dataset (ADFA-LD). Training time is reduced by implementing CRR in the place of normal LSTM. Normal call sequence training is given to the model. In that approach, the next integer was predicted through training provided probability distribution. That proposed method had a lower convergence speed. Hence, Robin Gassais et al. [53] proposed host-based automated IDS by combining machine learning (ML) algorithms and tracing techniques to improve the convergence speed. That proposed approach used Random Forest (FT) Gradient Boosted Trees (GBT) for intrusion detection. Information on tracing was obtained from user space and kernel space. Adaptation based tuning for new devices are explained in that approach. The system has shown better accuracy in detecting threats and alerting the system. In that proposed approach, the network was analyzed only in series. Thus, it does not provide higher detection accuracy.

By focusing on higher network accuracy, Prachi Deshpande et al. [54] proposed HBIDS by analyzing system call traces and alerting users if any threats were detected. K-nearest neighbor (kNN) was used as a classifier for tracing. The kNN allows easy inclusion of new training data. The network was analyzed through the frequency of failed system calls over successful system calls. The host-based IDS methods and its outcome are illustrated in Table V.

TABLE V. HOST BASED IDS METHODS

Author	Method	Dataset	Outcome	Advantages	Drawbacks
Chawla et al. [52]	Convolutional Neural Network Gated Recurrent Unit	Australian Defence Force Academy Linux Dataset	TDR - 100% FAR - 60%.	More accurate with reduced training time.	Slow convergence speed
Robin Gassais et al. [53]	Random Forest Gradient Boosted Trees	User Space Kernel Space	DT- 1.23 RF- 1.68 GBT 9.28 SVM 6.79 MLP 2.16	Improvement in detection time.	Analyzing only in series, not parallel and need learning for detection
Prachi Deshpande et al. [54]	K-nearest neighbour Classifier	-	A- 96%	Improvement in Accuracy	Delayed Detection time

### C. Network Intrusion Detection Approaches

Since there are several methods were suggested by the authors to deal with intrusions into the network. Previous sections analyze host-based, specification-based, network-based and anomaly-based methods. Those methods do not perform well in detecting all kinds of attacks and have reduced detection accuracy. Moreover, the network intrusions are detected by clustering, hybrid and evolutionary algorithms reviewed in the upcoming section.

1) *Clustering based NIDS*: The clustering-based approaches are aimed at reducing network complexities and easing detection accuracy. Some of the clustering-based NIDS are listed below.

Luiz Fernando Carvalho et al. [55] proposed an unsupervised learning approach for NIDS for extracting features. A modified ant colony optimization algorithm was used to optimize the multi-dimensional flow of the network. That proposed method offers lower detection accuracy; hence, to overcome these drawbacks, Yanqing Yang et al. [56] proposed a fuzzy aggregation method using a modified density peak clustering algorithm (MDPCA) and deep belief networks (DBNs). In that proposed approach, the training sets were divided into sub-sets with the same number of attributes of various cluster centers. That proposed method enhances automatic feature selection.

Moreover, Wei Liang et al. [57] proposed a multi-feature data clustering optimization model to improve impersonation attacks' detection accuracy. Security coefficients and weighted distances were categorized based on priority thresholds in that proposed method. The similarity of multi-feature data was examined through distance metrics. That proposed algorithm halted the clustering while reaching the preset iteration or obtaining the best cluster.

2) *Evolutionary algorithm*: Evolutionary computing methods are also known as bio-inspired algorithms that are processed based on the behavior of biological organisms. Bio-

inspired algorithms are widely applied in many fields due to their simple processing. Evolutionary algorithm-based NIDS is given below.

Vajiheh Hajisalem and Shahram Babaie [58] proposed hybrid Artificial Bee Colony (ABC) and Artificial Fish Swarm (AFS) algorithms. In addition, the fuzzy C-means clustering and Correlation-based Feature Selection (CFS) techniques were adopted to avoid unnecessary features. The CART technique formulated the fuzzy rules for classifying normal and intrusion data in that approach. Further improving network intrusion classification, Chaouki Khammassi and Saoussen Krichen [59] proposed a genetic algorithm (GA) and logistic regression based learning algorithm. That proposed method was evaluated under the KDD99 dataset and the UNSW-NB15 dataset. In addition, three decision tree classifiers were adopted for performance evaluation. The performance measures examined for the KDD99 dataset are illustrated in the table.

3) *Classification based NIDS*: Classification-based approaches are adopted for classifying malicious and normal behavior of the network. The classifier is developed by training data and classifies network behavior in classification-based methods. The classification approaches are carried out in either a multi-class or single-class manner. Some of the classification approaches for NIDS are given below.

Jiyeon Kim et al. [60] proposed a CNN-based NIDS to examine the DOS attack evaluated on the KDD CUP 1999 dataset comprised of DOS, U2R, R2L and probing attacks. That proposed model detects attacks belonging to a similar category. Moreover, the CSE-CIC-IDS2018 is also used for examining advanced IDS attacks. That proposed method offers lower detection accuracy of 93%. To enhance detection accuracy for examining cyber-attacks, Guo Pu et al. [61] proposed a hybrid unsupervised cluster based NIDS by combining OC-SVM and subspace clustering methods. This proposed method was validated under the NSL-KDD dataset. The OCSVM was suitable for unlabelled data in which the data had a normal class only. That maps both the feature space with data under the kernel. The feature selection must be improved further for better performance.

Yang Jia et al. [62] proposed a new deep NN (NDNN) model for NIDS that comprised four hidden layers evaluated on KDD99 and NSL-KDD. In that model, the input, hidden, and output layers consist of 41 neurons, 100 neurons, and 5 neurons, respectively. The proposed model performs better in KDD99 datasets but cannot provide better detection accuracy. But the KDD99 provides better performance in different kinds of attacks. Ahmed Iqbal and Shabib Aftab [63] proposed a feed-forward and pattern recognition (PR) ANN model to improve detection accuracy. The ANN was trained by scaled conjugate gradient, training functions and Bayesian regularization. That feed-forward network had multilayer neurons that were trained by Bayesian regularization. Then the PR was trained by a scaled conjugate gradient function. Arun Nagaraja et al. [64] proposed a UTTAMA classifier for detecting network intrusion; moreover, the Apriori algorithm based Frequent Pattern (FP) max algorithm was used. That

proposed classifier was tested under KDD-41 and KDD-19 datasets.

4) *Hybrid approaches*: Hybrid approaches are developed by adding various classification algorithms, methods and techniques. Some of the hybrid approaches are listed below.

Ansam Khraisat et al. [65] proposed a hybrid C5 decision tree classifier and OC-SVM, estimated on network security laboratory-knowledge discovery in a database (NSL-KDD). The proposed model used the attributes of benign samples, which does not use the data from other samples. The OCSVM classifier converts instances into a high dimensional attributes space and locates a suitable boundary hyperplane. That approach can detect the intrusions with only limited samples whereas, it has lower accuracy. Muhammad Ashfaq Khan [66] proposed a hybrid convolutional recurrent neural network-based NIDS. That method was estimated on the CSE-CIC-IDS2018 dataset that comprised seven types of attacks. Moreover, the network was mixed with traffic and non-traffic nodes to examine the suggested method. That suggested method was only investigated on the single dataset.

Different NIDS approaches are illustrated in Table VI.

TABLE VI. DIFFERENT APPROACHES FOR NETWORK INTRUSION DETECTION

Author	Method	Performance measures	Advantages	Drawbacks
Luiz Fernando Carvalho et al. [55]	Ant colony optimized digital signature	TPR is 93% for 1% FPR	Suitable for anomaly detection in large scale network	Lower detection accuracy
Yanqing Yang et al. [56]	MDPCA and DBN	A-90.21% DR-96.22% FPR-17.15%	Reduces the complexities in training sub-sets	Cannot detect the R2L and U2L attacks
Wei Liang et al. [57]	Multi-feature data clustering	DA-0.95 for 75 features	Detection accuracy is improved during a high overlap	Only detect the known attacks
Vajiheh Hajisalem and Shahram Babaie [58]	CFS techniques	FPR-0.01% DR-99%	Lower overhead	Need to reduce network complexity
Chaouki Khammassi and Saoussen Krichen [59]	genetic algorithm (GA) and logistic regression based learning algorithm	FAR-99.81% DR-1.105%	Better accuracy in detecting DOS attack	Need to reduce the misclassified results
Jiyeon Kim et al. [60]	CNN	P-1 A-93% R-1 F-1	New attacks are identified through the CSE-	Only detect the attack belongs to the same category

			CIC-IDS2018 dataset	
Guo Pu et al. [61]	SSC-OCSVM	DR-1 @ 0.05 false alarm rate	Detect large fraction of attacks in lower FAR	Lower detection accuracy
Yang Jia et al. [62]	NDNN	A-0.999 F-0.9998 R-0.9997	KDD99 dataset provides the best performance in all kinds of attacks detection	More network simulation experiments are needed
Ahmed Iqbal and Shabib Aftab [63]	FFANN	A-99.8356% MSE-0.0050	KDD CUP 99 examines different attacks	More improvements needed
Arun Nagaraja et al. [64]	UTTAMA	A-99.952%	A new membership function achieves a dimensional reduction	New distance functions are needed to improve performance efficacy.
Ansam Khraisat et al. [65]	OCSVM	A-83.24%	Detect the intrusions with fewer samples	Lower accuracy for detecting different tasks
Muhammad Ashfaq Khan [66]	HCRNNIDS	A-97.75% P-0.9633 R-0.9712 F-0.976 DR-0.97 FAR-2.5	Provides sufficient security against malicious nodes	The proposed method is tested on a single dataset only

5) *Discussion on traditional methods*

- The anomaly-based methods follow some detection rules based on some considerations thus, it is not capable of detecting all faults in the network. Moreover, the rule-defining process is degraded by the protocols, whereas the custom protocols also affect the rule-defining process.
- The clustering-based NIDS methods reduce the computational complexities in the network by grouping large datasets. In contrast, these methods affect detection rates in positive and negative ways. Moreover, the clustering approaches consume more time and can be affected by local minima.
- Classification-based NIDS approaches to improve the detection accuracy of network intrusion. Due to its adaptive nature, more data can be trained and tested using the ANN. On the other hand, the detection performance is improved by adding more layers within the network. Thus resource utilization is increased, resulting in over-fitting issues. Furthermore, that approaches needs relevant information for detecting unknown attacks.
- In evolutionary algorithm based NIDS, the detection accuracy and intrusion detection performance depend

on the algorithm used. The meta-heuristic algorithms adopt the behavior of living organism; thus, it does not require prior information about the network. The evolutionary algorithms do not affect by network noises. At the same time, the evolutionary algorithms are hard to map with network related issues.

- Compared with other methods, hybrid NIDS approaches have better performance but require high resources due to the combination of multiple techniques.

From the above analysis, it was noted that the rule-based approaches are affected by the protocols and clustering approaches increases the complexity due to large datasets and the classification based approaches needs more layers to provide an accurate results.

#### D. Deep Learning (DL) and Machine Learning based NIDS

Several methods are reviewed in the literature for detecting intrusion in the network, but they lack data management and feature learning. Both are considered important issues in IDS. Thus the DL based approaches are developed for detecting security threats in networks. Some of the DL based approaches are listed below.

Quamar Niyaz et al. [67] proposed a DL approach for detecting intrusion in the network in which self-taught learning (STL) was adopted. That proposed STL method comprised two stages wherein the representation of good features was learnt from unbalanced data collection. Then that learn representation was employed for labelled data that was adopted for classification. In that approach, the NSL-KDD dataset was adopted, which was minimized version of the KDD Cup 99 dataset. The sparse auto-encoder and soft-max regression matrix-based IDS were implemented in that paper. Furthermore, the performance of the IDS was enhanced by adding an extension of auto-encoders and stacked auto-encoder.

Hongpo Zhang et al. [68] proposed a denoising auto-encoder (DAE) and a weight loss function to select appropriate features for detecting intrusion by minimizing the feature dimensionality. Then, the chosen features were categorized through the multilayer perceptron (MLP) to identify the intrusion. In that proposed method, one key feature selection adds the weights to several samples' loss functions, which eases the feature selection. In that method, 12 features were chosen among 202 features with a 5.9% selection ratio. In MLP, two hidden layers were used for classification, which yielded an accuracy of 98.80%.

Fahimeh Farahnakian and Jukka Heikkonen [69] proposed a deep auto-encoder (DAE)-based method to detect network attacks caused by several vulnerabilities. That proposed DAE was trained in a greedy layer-wise fashion to evade the over-fitting and local optima. That proposed approach was executed through training and testing phases. During the training process, the system did the adoption of the training dataset and the creation of the DAE model. During the testing phase, a model was used for labelling test data. The input layer denotes 117 features of the dataset, and then the hidden layer chooses

32 features. In that method, the sigmoid function was chosen for hidden layers.

Vinayakumar et al. [70] proposed a deep NN (DNN) for detecting intrusion and classifying cyber-attacks. The Apache Spark cluster computing platform designed the scalable computing architecture in that proposed method. That proposed architecture explores the processing competence of a general purpose graphical processing unit (GPGPU) for rapid network investigation. The proposed architecture of DNN comprises five hidden layers for extracting complex features and pattern recognition ability. That proposed method was executed for different classifiers such as KDDCup 99, NSL-KDD, UNSW NB-15 and WSN-DS. Moreover, the result implies that the dataset of KDDCup99 and NSLKDD offer higher accuracy in the 95% to 99% range. The performance measures for binary classification of the KDDCup99 dataset with five layers of DNN are illustrated in the table.

Faten Louati and Farah Barika [71] proposed a DL based multi-agent system (MAS) for detection by combining the multi-agent features with DL algorithms. In that approach, the agents were generated by three algorithms: MLP, auto-encoder and k-nearest neighbour in which the autoencoder was used for feature reduction. In contrast, MLP and autoencoder were used for classification. In that approach, the MLP was a subset of DNN that was based on the backpropagation (BP) algorithm. Moreover, in autoencoders, the output layer had a similar number of nodes as the input layer. The auto encoders minimize the number of features from 120 to 10. In addition, the MAS enhances the network performance through proactivity and reactivity, which eases intrusion detection. The MAS comprises a pre-processor agent, reducer agent, classifier agent and decision-maker agent.

Soosan Naderi Mighan and Mohsen Kahani [72] proposed a hybrid DL and support vector machine (SVM) for feature extraction and classification. The stacked autoencoder (SAE) was used to reduce feature sets' dimensional reduction. In addition, the SVM was adopted for classification. In the pre-processing stage, symbolic features were converted into numeric values that range from zero to several symbols. The data normalization step was executed to minimize the dimension for all attributes. Then the latent features were extracted in the second phase; after that, the DL approach detected the attack. That suggested model was investigated on the ISCX IDS UNB dataset and it results in faster execution time.

Kasongo and Yanxia Sun [73] proposed feed-forward DNN (FFDNN) along with wrapper based feature selection (WFEU). The proposed method used an extra tree algorithm for optimum feature selection. That proposed method was evaluated under AWID and UNSW-NB15 datasets. The UNSW-NB15 dataset comprised 39 numerical features, three input nominal features had the feature type of binary, float and integer. In the initial phase, the FFDNN was implemented; after that, the WFEU was included. The experimental results demonstrate that the wrapper-based feature extraction was effective for the UNSW-NB15 dataset. In AWID, the

proposed model was implemented with a whole set of features; after that, 26 features were used from 154 attributes.

Sandeep Gurung et al. [74] proposed a DL and sparse auto-encoders for feature learning. That proposed method was trained by the NSL-KDD dataset that yields the output value as 0 (normal user) or 1 (intruder). In a pre-processing step, the numeric parameters were replaced instead of non-numeric parameters then the data was normalized. That proposed method reduced the false alarm rate lower than the signature-based method. Mike Nkongolo et al. [75] proposed a novel dataset, UGRansome1819, to detect unknown network attacks like zero-day threats. That proposed dataset benefited from unknown attacks that were not explored before and could not be observed by known attacks that were more efficient than the KDD99 and NSL-KDD datasets.

By analyzing several DL-based intrusion detection approaches, it has been concluded that deep network models examine all types of intrusion in the network. Furthermore, the deep networks can classify the intrusion type thus, it can be useful for detecting unknown attacks in the network. The node with an Internet protocol (IP) address and Internet produces network traffic that blocks the service to users [76]. However, the ML based methods are applicable in ad-click prognosis systems, and a two-way authentication system provides a secure connection between digital environments [77, 78]. If the attack arises from several distributed hosts, it will cause distributed denial of service attack (DDoS) that is more harmful than normal DoS [79, 80]. Illegitimate users create the DDoS to deny the server provided services [81]. Moreover, the system performance is improved by adding auto-encoders that increase detection accuracy.

The DL-based IDS with its pros and cons are illustrated in Table VII.

TABLE VII. DL BASED METHODS FOR IDS

Author	Method	Datasets	Classifier	Outcome	Advantages	Disadvantages
Quamar Niyaz et al. in 2016 [67]	DL	NSL-KDD dataset	Auto encoder	-	Improved performance by adding NB-tree and random-tree classifiers	The proposed method was not evaluated under real time IDS
Hongpo Zhang et al. in 2018 [68]	denoising auto-encoder (DAE)	UNSW-NB dataset	MLP	A-98.80% F-0.952 P-95.98% R-94.43% FPR-0.57%	Suitable for high speed network	-
Fahimeh Farahnakian and Jukka Heikkonen in	DAE	KDD-CUP 99	Softmax	A-94.71% FA-0.42% DR-95.53%	A greedy unsupervised layer-wise	The imposition of sparsity constraints

2018 [69]						training mechanisms improve performance	autoencoders is not illustrated in the paper.
Vinayakumar et al. in 2019 [70]	DNN	KDD-CUP 99, NSL-KDD and UNSW NB-15, WSN-DS	DT, AB, RF, LR, NB, KNN, SVM-rbf	A-0.927 P-0.994 R-0.915 F-0.953 (KDDC up99)		Better performance than machine learning classifiers	Failed to monitor DNS and BGP
Faten Louati and Farah Barika in 2020 [71]	DL-MAS	KDDC up99		A-99.95% P-00.32% FPR-0.17% FNR-0.68% TNR-99.83% TPR-99.32% DR-99.32%		The system tolerates the failure of one or more agents	Lower mobility and cloning
Soosan Naderi Mighan and Mohsen Kahani in 2018 [72]	DL-SAE	ISCX IDS UNB dataset	DL-SVM	A-0.902 TPR-0.903 TNR-0.902 P-0.903 R-0.903		Faster execution time	Not capable of managing large scale network traffic
Kasongo and Yanxia Sun in 2020 [73]	FFDNN	AWID, UNSW-NB15	SVM, decision tree (DT), random forest (RF), naive Bayes (NB) and k-nearest neighbour (kNN)	A-99.66%		Adapted for wireless application	Lower testing accuracy, 87.48% for binary classification
Sandeep Gurung et al. [74]	Sparsely AE	NSL-KDD dataset	Logistic classifier	A-87.2% P-84.7% R-92.8% Sp-80.7% NPV-90.7%		Used in any server which monitors the network activity	Lower accuracy

1) *Discussion on DL and ML approaches:* As mentioned earlier, the traditional methods have some limitations in intrusion detection; thus, the DL and ML based methods are used to resolve such issues. The insight on DL and ML based approaches are discussed below:

- The ML based methods provide better results in higher detection accuracy, but their performance completely relies on the data training.
- The DL methods efficiently examine enormous data in the network, but optimized layers improve their performance. Thus if the DL is used it must be used with the optimization algorithm to improve the computational efficacy.
- In a supervised learning based ML approach, classification is an important task; however, manual data labelling consumes more time for processing. Thus, the data labeling must be improved to get the benefits in ML approaches.
- The methods reviewed in the above section have used commonly used datasets such as KDD, CUP 99 etc., which do not contains the updated data of new types of attacks. While applying the methods to these datasets, the method's robustness cannot be determined. On the other hand, constructing a new dataset is expensive and needs expert knowledge.
- The better performance of IDS is not guaranteed while using ML with the dataset without real-world samples.

Findings: From these analysis, it is found that the method should be improved in terms of data collection, feature extraction and classification. Moreover, the inclusion of up-to-date dataset is also important factor to examine the unknown attacks. Moreover, these findings are more sufficient for proving efficient solution to improve the IDS against various attacks.

#### E. Limitations of NIDS

Although several methods are available for network intrusion detection, some limitations are listed below.

- Despite the ongoing research on IDS, intrusion detection techniques still have some drawbacks, such as slow detection time and high false detection. Due to the high detection time, it is hard to make working these methodologies for high speed networks.
- It is impossible to run soft-computing techniques on huge data with many features and imbalanced data. For analyzing such huge data, efficient sampling and feature selection techniques must be used.
- Parallely coordinated IDS are needed for large-scale and fast computing networks [82].
- Recently developed approaches perform better in detecting network intrusions but do not detect all types of attacks. Moreover, the available datasets comprised only a few attacks; thus, it is not effective for detecting all types of attacks. The up-to-date datasets may reveal new attacks in the network; thus, most of the recent

IDS cannot detect several types of attacks due to the unavailability of up-to-date datasets.

- The important issue with anomaly-based method is that it can examine the zero-day attack when properly modelled. The improper modelling of anomaly-based methods will raise the false alarm rate.
- The dataset must be integrated with the DL models to identify more attacks in the network to learn many patterns. In these cases, the dataset generation will be expensive.
- Most importantly, the unbalancing dataset may degrade network performance by reducing detection accuracy; thus, a balanced dataset is required for better dataset performance.
- The major challenge in IDS is the execution in the real-world environment, whereas most existing methods are not validated using a public dataset in the lab.
- Several methods suggested in existing works are complex in structure; this may cause extra overhead for the network process.
- To minimize the routing overhead of the network by proper feature selection, the researchers propose optimization algorithms, but the convergence speed of the existing algorithms will affects the feature extraction.
- The ML approach has some limitations, such as handling raw, high dimensional data and unlabeled data. Thus it cannot provide better classification in the presence of large datasets and complex data labelling. Thus the ML approaches are not suitable in the case of multi-classification function.
- The IDS are also adopted in IoT applications thus that deal with several sensor nodes in such methods. Not all kinds of IDS are suitable for this sense; only lightweight IDS is preferred in these applications due to minimum power utilization.
- IDS plays a major role in providing better system security than other systems. However, new malicious attacks are occurring in large amounts, making providing system security within computer networks a tedious task. Therefore, systematically updating available datasets would be the need of the hour.
- The challenges associated with IDS are false alarm rate, low detection rate, unbalanced datasets and response time. Misuse or signature based IDS are usually accompanied by some degree of false-positive alarm rates and are inefficient in detecting unknown or novel attacks. The main challenge of this system is updating the signatures of harmful intrusions. The challenges associated with anomaly based IDS are miscalculation in detection, lack of speed, difficulty in alerting and unbalanced datasheets.

## V. CONCLUSION

The IDS is a sufficient mechanism for securing the network against intrusions; thus, a comprehensive review of this topic is needed. This survey paper aimed to better understand intrusion detection in the network in accordance with different aspects. Different methods reviewed in this paper provide perceptive growth in intrusion detection. A wide range of methods was developed on this topic, each with advantages and drawbacks. A comparative analysis of different methods is given in this paper. Initially, several attacks occur in the networks were analyzed. After that, the NIDS is categorized under the detection technique and monitor platform. Further, the detection techniques are specification-based, anomaly-based and hybrid methods. Similarly, depending on the monitoring platform, the monitoring platform is classified as network-based or host-based. Additionally, classification-based, algorithm-based, clustering-based IDS are reviewed in detail. Tables provided deep insight into these methods and their pros and cons in each section. Different datasets are used in several methods to show the performance efficacy of the suggested method. It is found that most existing approaches detect few attacks only due to the availability of up-to-date datasets. By analyzing different methods, IDS is limited; the signature-based IDS are not effective for detecting the type of attack; thus, efficient IDS are modelled for detecting several kinds of attacks on the network.

The anomaly-based IDS effectively detect different types of attacks in the network, resulting in a higher false alarm rate; hence in future work, the anomaly-based approach will be improved by minimizing the false alarm rate. Even though the anomaly-based approaches can detect the zero-day attack, they cannot provide the desired outcome if it is not properly designed. To provide these issues, the ML and DL approaches are developed. The important task in intrusion detection is feature extraction. In the view of feature extraction, the DL outperforms ML by automatically extracting features. In addition, the survey shows that the classification-based methods had better performance in intrusion detection than all other methods. In addition, DL models are more effective for classification. Thus, it is concluded that DL approaches are more effective for detecting unknown attacks in the network. Most DL-based approaches are validated through different datasets in which the detection accuracy is based on feature selection. For this purpose, optimization algorithms are adopted. But single algorithms are not improving optimum feature selection. Since no methods were developed for unknown attack detection in the network with optimized DL. On the other hand, it is observed that the lack of an up-to-date dataset degrades intrusion detection in the network. This review cannot affirm the best method to detect unknown attacks in the network, but it suggests an effective way to detect unknown attacks. The future research direction of unknown attack detection is deliberated in the next section.

### A. Future Works

The future scope of IDS is listed below.

This review verified that DL approaches are more effective, but their performance can be improved by adding an

efficient algorithm. Hence, hybrid DL approaches are suggested for detecting known and unknown attacks in future work. In order to improve the detection of unknown attacks, it is important to use the up-to-date method. Future work will focus on unknown attack detection in the network; hence, the recently developed UGRansome1819 dataset is suggested. The intrusion detection not only improved by the dataset but also improved by adding an efficient feature section approach. In fact, hybrid pattern search whale optimization algorithm will be effective for optimal feature selection. Pattern search is a non-derivative algorithm that is suitable for updating an optimum weight as well as the whale optimization algorithm (WOA) has a better convergence speed. Hence adopting both these algorithms will improve the feature selection. The important task after feature selection is classification, which provides the final outcome about the attacks in the network. At the same time, the lack of processing speed will reduce the detection accuracy. Furthermore, the hybrid bi-directional long short term memory (Bi-LSTM) with the gated recurrent unit (GRU) has to be implemented to classify unknown attacks in the network. These improved feature extraction and classification approaches will be effective solutions for detecting unknown attacks in the network.

## REFERENCES

- [1] M. Ozkan-Okay, R. Samet, Ö. Aslan, & D. Gupta, "A Comprehensive Systematic Literature Review on Intrusion Detection Systems," IEEE Access, 2021.
- [2] R.V. Mendonça, A.A. Teodoro, R.L. Rosa, M. Saadi, D.C. Melgarejo, P.H. Nardelli, & D.Z. Rodríguez, "Intrusion detection system based on fast hierarchical deep convolutional neural network," IEEE Access, vol. 9, pp. 61024-61034, 2021.
- [3] M. Al-Qatf, Y. Lasheng, M. Al-Habib, & K. Al-Sabahi, "Deep learning approach combining sparse autoencoder with SVM for network intrusion detection," Ieee Access, vol. 6, pp. 52843-52856, 2018.
- [4] A. Borkar, A. Donode, & A. Kumari, "A survey on Intrusion Detection System (IDS) and Internal Intrusion Detection and protection system (IIDPS)," In 2017 International conference on inventive computing and informatics (ICICI), IEEE, pp. 949-953, 2017.
- [5] W. Li, S. Tug, W. Meng, & Y. Wang, "Designing collaborative blockchained signature-based intrusion detection in IoT environments," Future Generation Computer Systems, vol. 96, pp. 481-489, 2019.
- [6] M. Torabi, N.I. Udzir, M.T. Abdullah, & R. Yaakob, "A review on feature selection and ensemble techniques for intrusion detection system," network, vol. 1, pp. 2, 2021.
- [7] M. Uğurlu, and İ.A. Doğru, "A survey on deep learning based intrusion detection system," In 2019 4th International Conference on Computer Science and Engineering (UBMK), IEEE pp. 223-228, 2019.
- [8] M. Almseidin, M. Alzubi, S. Kovacs, & M. Alkasassbeh, "Evaluation of machine learning algorithms for intrusion detection system," In 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), IEEE, pp. 000277-000282, 2017.
- [9] M.Y. AlYousef, & N.T. Abdelmajeed, "Dynamically Detecting Security Threats and Updating a Signature-Based Intrusion Detection System's Database," Procedia Computer Science, vol. 159, pp. 1507-1516, 2019.
- [10] M. Dua, "Machine learning approach to IDS: A comprehensive review," In 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE pp. 117-121, 2019.
- [11] M. Maithem, and G.A. Al-sultany, "Network intrusion detection system using deep neural networks," In Journal of Physics: Conference Series, IOP Publishing vol. 1804, no. 1, pp. 012138, 2021.
- [12] G. Karatas, O. Demir, and O.K. Sahingoz, "Deep learning in intrusion detection systems, In 2018 International Congress on Big Data," Deep

- Learning and Fighting Cyber Terrorism (IBIGDELFT), IEEE pp. 113-116, 2018.
- [13] O.M.A. Alsayibani, E. Utami, and A.D. Hartanto, "Survey on Deep Learning Based Intrusion Detection System," *Telematika*, vol. 14, no. 2, pp. 86-100, 2021.
- [14] A.R. Khan, M. Kashif, R.H. Jhaveri, R. Raut, T. Saba, and S.A. Bahaj, "Deep learning for intrusion detection and security of Internet of things (IoT): current analysis, challenges, and possible solutions," *Security and Communication Networks*, vol. 2022, 2022.
- [15] Y. Li, R. Liu, X. Liu, H. Li, & Q. Sun, "Research on information security risk analysis and prevention technology of network communication based on cloud computing algorithm," In *Journal of Physics: Conference Series*, IOP Publishing vol. 1982, no. 1, pp. 012129, 2021.
- [16] T.H. Hadi, "Types of Attacks in Wireless Communication Networks," *Webology*, vol. 19, no. 1, 2022.
- [17] M.V Pawar, & J. Anuradha, "Network security and types of attacks in network," *Procedia Computer Science*, vol. 48, pp. 503-506, 2015.
- [18] P. Amish, & V.B. Vaghela, "Detection and prevention of wormhole attack in wireless sensor network using AOMDV protocol," *Procedia computer science*, vol. 79, pp. 700-707, 2016.
- [19] A.U. Rehman, S.U. Rehman, & H. Raheem, "Sinkhole attacks in wireless sensor networks: a survey," *Wireless Personal Communications*, vol. 106, no. 4, pp. 2291-2313, 2019.
- [20] Z. Gavric, & D. Simic, "Overview of DOS attacks on wireless sensor networks and experimental results for simulation of interference attacks," *Ingeniería e Investigación*, vol. 38, no. 1, pp. 130-138, 2018.
- [21] B. Bhushan, & G. Sahoo, "Recent advances in attacks, technical challenges, vulnerabilities and their countermeasures in wireless sensor networks," *Wireless Personal Communications*, vol. 98, no. 2, pp. 2037-2077, 2018.
- [22] M. Conti, N. Dragoni, and V. Lesyk, "A survey of man in the middle attacks," *IEEE communications surveys & tutorials*, vol. 18, no. 3, pp. 2027-2051, 2016.
- [23] K.L. Chiew, K.S.C. Yong, and C.L. Tan, "A survey of phishing attacks: Their types, vectors and technical approaches," *Expert Systems with Applications*, vol. 106, pp. 1-20, 2018.
- [24] A. Thakkar, & R. Lohiya, "A survey on intrusion detection system: feature selection, model, performance measures, application perspective, challenges, and future research directions," *Artificial Intelligence Review*, pp. 1-111, 2021.
- [25] A. Khraisat, I. Gondal, P. Vamplew, & J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," *Cybersecurity*, vol. 2, no. 1, pp. 1-22, 2019.
- [26] M. Almansor, & K.B. Gan, "Intrusion detection systems: principles and perspectives," *Journal of Multidisciplinary Engineering Science Studies*, vol. 4, no. 11, pp. 2458-2925, 2018.
- [27] A. Adnan, A. Muhammed, A.A Abd Ghani, A. Abdullah, & F. Hakim, "An Intrusion Detection System for the Internet of Things Based on Machine Learning: Review and Challenges," *Symmetry*, vol. 13, no. 6, pp. 1011, 2021.
- [28] M. Masdari, and H. Khezri, "A survey and taxonomy of the fuzzy signature-based intrusion detection systems," *Applied Soft Computing*, vol. 92, pp. 106301, 2020.
- [29] A. Lazarevic, V. Kumar, & J. Srivastava, "Intrusion detection: A survey," In *Managing cyber threats*, Springer, Boston, MA pp. 19-78, 2005.
- [30] R.A. Beyah, M.C. Holloway, & J.A. Copeland, "Invisible trojan: an architecture, implementation and detection method," In *The 2002 45th Midwest Symposium on Circuits and Systems*, 2002. MWSCAS-2002., IEEE, vol. 3, pp. III-III, 2002.
- [31] S. Zavrak, & M. İskefiyeli, "Anomaly-based intrusion detection from network flow features using variational autoencoder," *IEEE Access*, vol. 8, pp. 108346-108358, 2020.
- [32] S. Dwivedi, M. Vardhan, S. Tripathi, & A.K. Shukla, "Implementation of adaptive scheme in evolutionary technique for anomaly-based intrusion detection," *Evolutionary Intelligence*, vol. 13, no. 1, pp. 103-117, 2020.
- [33] B.A. Tama, M. Comuzzi, & K.H. Rhee, "TSE-IDS: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system," *IEEE Access*, vol. 7, no. 94497-94507, 2019.
- [34] N.T. Van, & T.N. Thinh, "An anomaly-based network intrusion detection system using deep learning," In *2017 international conference on system science and engineering (ICSSE)*, IEEE, pp. 210-214, 2017.
- [35] R. Kumar, & D. Sharma, "HyINT: signature-anomaly intrusion detection system, In *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, pp. 1-7, 2018.
- [36] A. Le, J. Loo, K.K. Chai, & M. Aiash, "A specification-based IDS for detecting attacks on RPL-based network topology," *information*, vol. 7, no. 2, pp. 25, 2016.
- [37] H. Esquivel-Vargas, M. Caselli, & A. Peter, "Automatic deployment of specification-based intrusion detection in the BACnet protocol," In *Proceedings of the 2017 Workshop on Cyber-Physical Systems Security and Privacy*, pp. 25-36, 2017.
- [38] A. Althubaity, H. Ji, T. Gong, M. Nixon, R. Ammar, & S. Han, "ARM: A hybrid specification-based intrusion detection system for rank attacks in 6TiSCH networks," In *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, IEEE, pp. 1-8, 2017.
- [39] S. Einy, C. Oz, & Y.D. Navaei, "The anomaly-and signature-based IDS for network security using hybrid inference systems," *Mathematical Problems in Engineering*, 2021.
- [40] M. Mehmood, T. Javed, J. Nebhen, S. Abbas, R. Abid, G.R. Bojja, & M. Rizwan, "A hybrid approach for network intrusion detection," *CMC-Comput. Mater. Contin.*, vol. 70, pp. 91-107, 2022.
- [41] B. Subba, S. Biswas, & S. Karmakar, "A neural network based system for intrusion detection and attack classification," In *2016 Twenty Second National Conference on Communication (NCC)*, IEEE, pp. 1-6, 2016.
- [42] N. Ádám, B. Madoš, A. Baláz, & T. Pavlik, "Artificial neural network based IDS," In *2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMII)*, IEEE, pp. 000159-000164, 2017.
- [43] V.D. Mane, & S. Pawar, "Anomaly based ids using backpropagation neural network," *International Journal of Computer Applications*, vol. 136, no. 10, pp. 29-34, 2016.
- [44] H. Gharace, & H. Hosseinvand, "A new feature selection IDS based on genetic algorithm and SVM," In *2016 8th International Symposium on Telecommunications (IST)*, IEEE, pp. 139-144, 2016.
- [45] K. Zhang, F. Zhao, S. Luo, Y. Xin, & H. Zhu, "An intrusion action-based IDS alert correlation analysis and prediction framework," *IEEE Access*, vol. 7, pp. 150540-150551, 2019.
- [46] B. Subba, S. Biswas, & S. Karmakar, "False alarm reduction in signature-based IDS: game theory approach," *Security and Communication Networks*, vol. 9, no. 18, pp. 4863-4881, 2016.
- [47] M.A. Ambusaidi, X. He, P. Nanda, & Z. Tan, "Building an intrusion detection system using a filter-based feature selection algorithm," *IEEE transactions on computers*, vol. 65, no. 10, pp. 2986-2998, 2016.
- [48] S. Kumar, A. Viinikainen, & T. Hamalainen, "Machine learning classification model for network based intrusion detection system," In *2016 11th International Conference for Internet Technology and Secured Transactions (ICITST)*, IEEE, pp. 242-249, 2016.
- [49] H. Larijani, J. Ahmad, & N. Mtetwa, "A novel random neural network based approach for intrusion detection systems," In *2018 10th Computer Science and Electronic Engineering (CEECE)*, IEEE, pp. 50-55, 2018.
- [50] S. Nayyar, S. Arora, & M. Singh, "Recurrent neural network based intrusion detection system," In *2020 International Conference on Communication and Signal Processing (ICCCSP)*, IEEE. pp. 0136-0140, 2020.
- [51] N. Oliveira, I. Praça, E. Maia, & O. Sousa, "Intelligent cyber-attack detection and classification for network-based intrusion detection systems," *Applied Sciences*, vol. 11, no. 4, pp. 1674, 2021.
- [52] A. Chawla, B. Lee, S. Fallon, & P. Jacob, "Host based intrusion detection system with combined CNN/RNN model," In *Joint European*



- Conference on Machine Learning and Knowledge Discovery in Databases, pp. 149-158, 2018. Springer, Cham.
- [53] R. Gassais, N. Ezzati-Jivan, J.M. Fernandez, D. Aloise, & M.R. Dagenais, "Multi-level host-based intrusion detection system for Internet of things," *Journal of Cloud Computing*, vol. 9, no. 1, pp. 1-16, 2020.
- [54] P. Deshpande, S.C. Sharma, S.K. Peddoju, & S. Junaid, "HIDS: A host based intrusion detection system for cloud computing environment," *International Journal of System Assurance Engineering and Management*, vol. 9, no. 3, pp. 567-576, 2018.
- [55] L.F. Carvalho, S. Barbon Jr, L. de Souza Mendes, & M.L. Proenca Jr, "Unsupervised learning clustering and self-organized agents applied to help network management," *Expert Systems with Applications*, vol. 54, pp. 29-47, 2016.
- [56] Y. Yang, K. Zheng, C. Wu, X. Niu, & Y. Yang, "Building an effective intrusion detection system using the modified density peak clustering algorithm and deep belief networks," *Applied Sciences*, vol. 9, no. 2, pp. 238, 2019.
- [57] W. Liang, K.C. Li, J. Long, X. Kui, & A.Y. Zomaya, "An industrial network intrusion detection algorithm based on multifeature data clustering optimization model," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2063-2071, 2019.
- [58] V. Hajisalem, & S. Babaie, "A hybrid intrusion detection system based on ABC-AFS algorithm for misuse and anomaly detection," *Computer Networks*, vol. 136, pp. 37-50, 2018.
- [59] C. Khammassi, & S. Krichen, "A GA-LR wrapper approach for feature selection in network intrusion detection," *computers & security*, vol. 70, pp. 255-277, 2017.
- [60] J. Kim, J. Kim, H. Kim, M. Shim, & E. Choi, "CNN-based network intrusion detection against denial-of-service attacks," *Electronics*, vol. 9, no. 6, pp. 916, 2020.
- [61] G. Pu, L. Wang, J. Shen, & F. Dong, "A hybrid unsupervised clustering-based anomaly detection method," *Tsinghua Science and Technology*, vol. 26, no. 2, pp. 146-153, 2020.
- [62] Y. Jia, M. Wang, & Y. Wang, "Network intrusion detection algorithm based on deep neural network," *IET Information Security*, vol. 13, no. 1, pp. 48-53, 2019.
- [63] A. Iqbal, & S. Aftab, "A Feed-Forward and Pattern Recognition ANN Model for Network Intrusion Detection," *International Journal of Computer Network & Information Security*, vol. 11, no. 4, 2019.
- [64] A. Nagaraja, & B. Uma, "UTTAMA: An intrusion detection system based on feature clustering and feature transformation," *Foundations of Science*, vol. 25, no. 4, pp. 1049-1075, 2020.
- [65] A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, & A. Alazab, "Hybrid intrusion detection system based on the stacking ensemble of c5 decision tree classifier and one class support vector machine," *Electronics*, vol. 9, no. 1, pp. 173, 2020.
- [66] M.A. Khan, "HCRNNIDS: hybrid convolutional recurrent neural network-based network intrusion detection system," *Processes*, vol. 9, no. 5, pp. 834, 2021.
- [67] A. Javaid, Q. Niyaz, W. Sun, & M. Alam, "A deep learning approach for network intrusion detection system," *Eai Endorsed Transactions on Security and Safety*, vol. 3, no. 9, pp. e2, 2016.
- [68] H. Zhang, C.Q. Wu, S. Gao, Z. Wang, Y. Xu, & Y. Liu, "An effective deep learning based scheme for network intrusion detection, In 2018 24th International Conference on Pattern Recognition (ICPR)," IEEE, pp. 682-687, 2018.
- [69] F. Farahnakian, & J. Heikkonen, "A deep auto-encoder based approach for intrusion detection system, In 2018 20th International Conference on Advanced Communication Technology (ICACT)," IEEE, pp. 178-183, 2018.
- [70] R. Vinayakumar, M. Alazab, K.P. Soman, P. Poornachandran, A. Al-Nemrat, & S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41525-41550, 2019.
- [71] F. Louati, & F.B. Ktata, "A deep learning-based multi-agent system for intrusion detection," *SN Applied Sciences*, vol. 2, no. 4, pp. 1-13, 2020.
- [72] S.N. Mighan, & M. Kahani, "Deep learning based latent feature extraction for intrusion detection, In Electrical Engineering (ICEE)," Iranian Conference on, IEEE, pp. 1511-1516, 2018.
- [73] S.M. Kasongo, & Y. Sun, "A deep learning method with wrapper based feature extraction for wireless intrusion detection system," *Computers & Security*, vol. 92, pp. 101752, 2020.
- [74] S. Gurung, M.K. Ghose, & A. Subedi, "Deep learning approach on network intrusion detection system using NSL-KDD dataset," *International Journal of Computer Network and Information Security*, vol. 11, no. 3, pp. 8-14, 2019.
- [75] M. Nkongolo, J.P. van Deventer, & S.M. Kasongo, "UGRansome1819: A Novel Dataset for Anomaly Detection and Zero-Day Threats," *Information*, vol. 12, no. 10, pp. 405, 2021.
- [76] S. Shunmuganathan, R.D. Saravanan, & Y. Palanichamy, "Securing VPN from insider and outsider bandwidth flooding attack," *Microprocessors and Microsystems*, vol. 79, pp. 103279, 2020.
- [77] S. Saraswathi, V. Krishnamurthy, D.V.V. Prasad, R.K. Tarun, S. Abhinav, & D. Rushitaa, "Machine learning based Ad-click prediction system," *Int J Eng Adv Technol*, vol. 8, no. 6, pp. 3646-3648, 2019.
- [78] S. Shunmuganathan, "A Reliable Lightweight Two Factor Mutual Authenticated Session Key Agreement Protocol for Multi-Server Environment," *Wireless Personal Communications*, vol. 121, no. 4, pp. 2789-2822, 2021.
- [79] R.D. Saravanan, S. Loganathan, S. Shunmuganathan, & Y. Palanichamy, "Suspicious score based mechanism to protect web servers against application layer distributed denial of service attacks," *Int J Intell Eng Syst*, vol. 10, no. 4, pp. 147-156, 2017.
- [80] S.R. Devi, S. Saraswathi, & P. Yogesh, "A Cooperative Multilayer End-Point Approach to Mitigate DDoS Attack," *WSEAS Transactions on Information Science and Applications*, vol. 11, no. 1, pp. 1-11, 2014.
- [81] R. Saravanan, S. Shanmuganathan, & Y. Palanichamy, "Behavior-based detection of application layer distributed denial of service attacks during flash events," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 24, no. 2, pp. 510-523, 2016.
- [82] R. Singh, H. Kumar, R.K. Singla, & R.R. Ketti, "Internet attacks and intrusion detection system: A review of the literature," *Online Information Review*, 2017.

# Ọdigo Metaheuristic Optimization Algorithm for Computation of Real-Parameters and Engineering Design Optimization

Ikpo C V<sup>1</sup>, Akowuah E K<sup>2</sup>, Kponyo J J<sup>3</sup>, Boateng K O<sup>4</sup>

Department of Computer Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana<sup>1, 2, 4</sup>  
Department of Telecommunication Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana<sup>3</sup>

**Abstract**—This paper proposes a new population-based global optimization algorithm, Ọdigo Metaheuristic Optimization Algorithm–QMOA, for solving complex bounded-constraint/single objective real-parameter problems found in most engineering and scientific applications. It’s inspired by the human socio-cultural informal discipleship learning pattern inherent in the behavior of the Ndigbo peoples; the subject – primary (Nwa-ahja), in mercantile cycle grows to a secondary (Mazi) owing to the intuitive stratagem (dialect - Igba) embedded in an aged-long cultural model “Igba-ọsọ-ahja” (meaning, strategic marketing skills, and practice). The model mimics the search routine for satisfying a customer’s need in the market, built into exploration and exploitation applied in the mathematical model. About 30 complex classical unconstrained functions are tested, comparing results with that of five similar state-of-the-art algorithms. Also, 29 CEC-2017 single objective real constraint benchmark serious dimensional problems were simulated and compared against the winners of that competition. Validation includes statistical (t-test, p-value) comparison and for 50 Dimension constraint problems as QMOA demonstrated superior performance. TCS (9.18%), WBP (6.3%), PVDP (601%), RGP (319%), RBP (760%), GTCD (202%), HIMMELBLAU (4%), and CDP (88.12%) are the improvements made on 8 CEC-2020 engineering real design problems against the former best performances; QMOA is simple to implement, replicate and applicable across domains. Also, some new, improved optimum was obtained in Shubert and Schaffer 4 function compared to the global optimums.

**Keywords**—Human socio-cultural; nature-inspired; informal-learning; global optimization

## I. INTRODUCTION

Humans and animals face challenges within their time and space of habitation, and they attempt to solve the challenges by making decisions and selecting and combining variables influencing the conditions. The challenges range from simple to difficult-complex ones, but the satisfaction derived from attaining the goal motivates effort for solution pursuant [1]. Engineering has availed very good solutions for small scaled problems using exact methods, but such fails when the problem becomes special and high dimension, become very costly and time consuming [2]. Meanwhile, the study of nature showed complex problems solved by meta-ideas and heuristics. The aesthetics that describes the meta-heuristics provide solutions that are near-optimal yet scalable with problem dimensions [3] despite the difficult procedural

uncertainties [4]; the huge difficulty is associated with the mapping of routines called intelligence from rules or heuristics that describe events of nature which falls in a multidisciplinary field [5, 6]. Research in this direction has yielded several methodologies for solving engineering problems, yet more are anticipated [6]. This work aims to address some multidisciplinary domain concerns; a significant gap in balancing exploitation and exploration in populations of solution search impacts the state-of-art. Also, most recent works have scantily described the critical analogies of the metaphors that reflect the aesthetics of the target nature’s source with the derived mathematical models, while the majority favors hybridization. Also, only a handful of the existing algorithms had human behavior metaphors, which this work proposes. Based on life science, a simple category of existing solutions could be into biological and non-biological (abiotic) hybrids, Bio-Abiotic hybrids, Bio-Bio hybrids, and Abiotic – Abiotic hybrids; however other literature may use alternative categorizations such as Swarm, Evolutionary, and Human intelligence. Genetic algorithms (GA) led the natural biological methods [7]. Particle Swarm Optimization (PSO) is inspired by flocks of birds and schools of fish [8] [9]. A few others due to space constraints are; Artificial Bee Colony from bee foraging [10]; Ant Colony [11]. In literature, numerous applications of the metaheuristics includes scheduling, loading, packaging, design, and control [12], image processing, amongst numerous others. The abiotic category is based on artificial physical experiences, such as Tabu Search, which made use of the creation of a tabu list [13]; Water Evaporation Optimization (WEO), mimicking the evaporation of water [14, 15]; JAYA mimicking the gravitation towards success [15]; Atomic Orbital Search (AOS) [16], etc. Some modified/hybrids are; Grey Wolf and PSO [17] gave (GWO-PSO), MOGSABAT [18] from the multiobjective gravitational search algorithm, and the echolocation ability of the bat algorithm [19]. Many other metaheuristic methods can be found [20, 21]. QMOA is a new strategy proposed by this work; the data is from the human population shown in Section II. The aesthetics are based on informal learning. The mathematical relations are developed in Section III and experiments, results and discussions are also presented in Section III, while Section IV is the conclusion.

The data of this work is gathered from the Ndigbo people’s mercantilism. This ideology is found in major Market setups across the World, where Ndigbo are found in huge populations

[22, 23]. They cooperate and maintain this characteristic ideology they call “igba”; meaning stratagem. [24, 25]. The aesthetics; every male is disciple/given-chance-to-hands-on/learn informally in commerce backed by some form of agreement [26]; a model known to them as “Igba-oso-ahija” which means strategic marketing skills acquisition and practice. Ahija (market) is a solution space and holds all history of exploitations and explorations through Igba-oso-ahija model [27]. There exist huge risks and sacrifices, but the Ndigbo tolerates them [27].

A. QMOA Algorithm Description

In the Ahija environment, the ultimate is to become a Mazi; The initial population is generated randomly as Ahija-size. This is the “initialization mode”; The readiness, practicing, discipline, trading, cooperation, and the reluctance of the agents (known as umu-ahija in Igbo) is adjusted against the new environment each day; [from start-transduction mode to update-matching mode].

II. MODELING DATA AND AESTHETICS

The work started with a collection of data from a local ahija; the data is found at <https://data.mendeley.com/datasets/wt3vt72mph/1>. A few assumptions and facts extracted from data include but are not limited to the following parameters:

- 1) **NORMS:** (i) Every Mazi Own at least one shop. (ii) Every Nwa-ahija is attached to a Mazi, a shop, and an ahija. With an agreement, (iii) Death or risk are inevitable etc.
- 2) **AXUMES:** (i) Every Nwa-ahija must satisfy a certain set percent of discipleship requirements to become a Mazi.
- 3) **FACTS and Probable:** (i) Certain Nwa-ahija may succeed, fail, die, or get impeded. (ii) Certain Mazi may become greedy and unjust. (iii) Certain Umu-ahija had gotten second and third chances to make up, and many ahija exist.

B. Sample Size of Selected Market

Data in Table I shows a snapshot of the collection, and the values represent the sub-total in each case. For example, the column representing “Japan”; “JAPANLINE”; “shops:35”; Parts: [12: “Nissan”, 23: “Toyota Accessories”, NULL:” x”].

C. The Model - Ahija

The visualization of the setup of ahija as a system (inputs, process, and outputs) schematically looks like Fig. 1 (left, right)

Fig. 1(a) shows Ahija [n+9] described in Fig. 1(b) explicitly; the lines show the nonlinear relationships. The inner layers are shops and are associated with entities enlisted. The local Ahija are networked across major cities in Nigeria (Ibadan, Lagos, Onitsha, etc.) which affiliates to extensions in Countries like Japan and Germany. The Ahija primary agent (humans) are umu-ahija, and secondary are ndi-oso-ahija, Mazi, Bankers, customers (Regular and Non-Regular), suppliers, forwarding and clearing and etc. Meanwhile, the number of decision variables in sales, storage, borrowing etc., varies with constraints of environments like the cash flow, religion, and local/global politics etc.; taking shop 9 – GermanyLine Fig. 2; it comprises 1 –Mazi, 5 – Umu-ahija, 1 – Onye-oso-ahija, 12 –

Regular Customers, 100 – Emergency Customers and trading on Benz-Spare Parts as shown.

TABLE I. SAMPLED DATA FROM A MARKET AND SHOP DISTRIBUTIONS

Object	JAPAN	TOYOTA	GERMANY	AB	BAMENDA
Mazi	35	43	101	122	199
Umu-ahija	210	250	591	696	1162
Ndi-oso-ahija	35	43	101	122	199
Regular Customers	2204	2786	5039	19264	23616
Emergency Customers	2859	4646	10100	15595	20168
Shops					
Nissan and Toyota Parts	12	x	x	x	x
Toyota Accessories	23	x	x	x	x
Toyota-Spare-Parts	x	43	x	x	x
Benz-Spare-Parts	x	x	22	x	x
Audi Parts	x	x	31	x	x
Gold and Volkswagen	x	x	24	x	x
Benz Engine	x	x	11	x	x
Benz-Dashboard and Accessories	x	x	13	x	x
Cloth and Okrika	x	x	x	122	x
Tokunbo Fridges & Phones	x	x	x	x	199

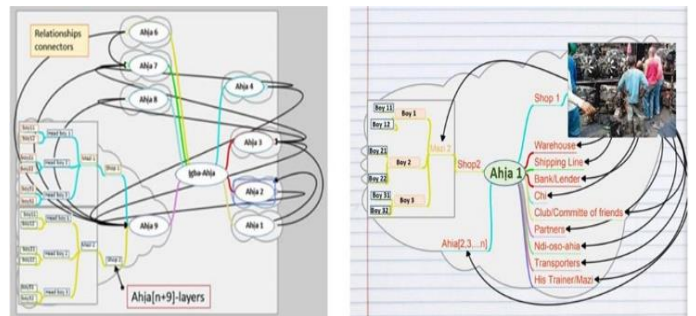


Fig. 1. Ahija business model and networks.

#	A/B	C	D	E	F	G	H	I	J	K	L
2	1/n Business Line	Mazi	Umu-ahija	Ndi-oso-ahija	Regular Customer	Emergency Customer	Good Type		Market Start Date (No of Shops)		
3	1	Germany Line - Shop 1	1	6	1	30	100 Benz-Spare Parts	Eke	10/1/21	1	
4	2	Germany Line - Shop 2	1	6	1	50	100 Benz-Spare Parts	Eke	10/1/21	1	
5	3	Germany Line - Shop 3	1	5	1	53	100 Benz-Spare Parts	Eke	10/1/21	1	
6	4	Germany Line - Shop 4	1	5	1	21	100 Benz-Spare Parts	Eke	10/1/21	2	
7	5	Germany Line - Shop 5	1	6	1	32	100 Benz-Spare Parts	Eke	10/1/21	1	
8	6	Germany Line - Shop 6	1	5	1	35	100 Benz-Spare Parts	Orle	10/2/21	1	
9	7	Germany Line - Shop 7	1	6	1	56	100 Benz-Spare Parts	Orle	10/2/21	1	
10	8	Germany Line - Shop 8	1	6	1	120	100 Benz-Spare Parts	Orle	10/2/21	1	
11	9	Germany Line - Shop 9	1	5	1	12	100 Benz-Spare Parts	Orle	10/2/21	1	
12	10	Germany Line - Shop 1	1	5	1	14	100 Benz-Spare Parts	Orle	10/2/21	1	
13	11	Germany Line - Shop 1	1	5	1	16	100 Benz-Spare Parts	Nkwo	10/4/21	1	
14	12	Germany Line - Shop 1	1	5	1	21	100 Benz-Spare Parts	Nkwo	10/4/21	1	
15	13	Germany Line - Shop 1	1	5	1	15	100 Benz-Spare Parts	Nkwo	10/4/21	1	
16	14	Germany Line - Shop 1	1	5	1	15	100 Benz-Spare Parts	Nkwo	10/4/21	1	
17	15	Germany Line - Shop 2	1	5	1	14	100 Benz-Spare Parts	Nkwo	10/4/21	1	
18	16	Germany Line - Shop 1	1	5	1	14	100 Benz-Spare Parts	Nkwo	10/4/21	1	
19	17	Germany Line - Shop 1	1	5	1	15	100 Benz-Spare Parts	Eke	10/5/21	1	
20	18	Germany Line - Shop 1	1	5	1	25	100 Benz-Spare Parts	Eke	10/5/21	1	
21	19	Germany Line - Shop 1	1	5	1	15	100 Benz-Spare Parts	Eke	10/5/21	1	
22	20	Germany Line - Shop 2	1	5	1	14	100 Benz-Spare Parts	Eke	10/5/21	1	
23	21	Germany Line - Shop 2	1	6	1	29	100 Benz-Spare Parts	Eke	10/5/21	1	
24	22	Germany Line - Shop 2	1	6	1	50	100 Benz-Spare Parts	Eke	10/6/21	1	
25	23	Germany Line - Shop 2	1	6	1	29	100 Benz-Spare Parts	Orle	10/6/21	1	
26	24	Germany Line - Shop 2	1	6	1	37	100 Audi Parts	Orle	10/6/21	1	

Fig. 2. Snap off idumota sampling.

In obtaining the adjacency list from the data, assumptions made included (1) (1/0 means connected/not-connected)

respectively; (2) also shop data is deterministic data at capture time, the network of the single shop nine(9) is modeled, and the simulation – Bayes graph is as shown

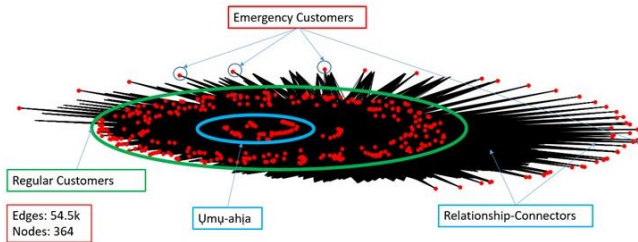


Fig. 3. The schematic representation of the network traffic.

The complex network of the shop (54500 edges, 364 nodes), on a market day named Orië; Fig. 3 is made. It depicts the intense cognitive field (energy) of nonlinear relationship maps responsible for transduced processes with experience (edges) of umu-ahja (nodes) that Orië day. Daily customer satisfaction time monotonically decreases with increasing edges, even with constraints in cycles. Beyond shop 9, thousands of shops contribute to the ahja data; the computer structural model is shown in Fig. 4.

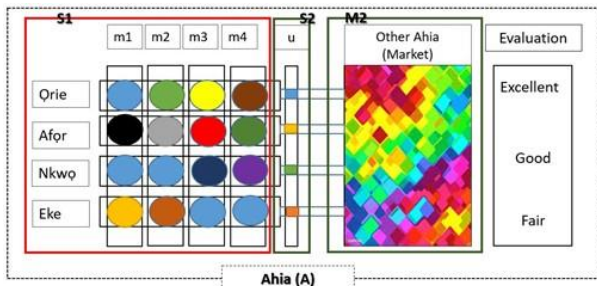


Fig. 4. Ahja environment across many shops.

S1, S2, m1, m2, m3, m4, and M2 represents: shops 1, ndi-oso-ahja, nwa-ahja-1, nwa-ahja-2, nwa-ahja-3, nwa-ahja-4, and many other distance ahja (markets). The updates are processes of the umu-ahja transforming on every market day (Orië, Afor, Nkwò, Eke). The colours is evidence uncertainty.

#### D. Initialization of Population (Market-Size)

OMOA; with the decision variables, Igba-oso-ahja and D-dimensions, the solution vector in the ahja can be represented as (1).

$$Mazi = [x_1, x_2, x_3, \dots, x_D] \quad (1)$$

The fitness value of each Mazi will be computed as a vector of (2);

$$f(Mazi) = f(x_1, x_2, x_3, \dots, x_D) \quad (1)$$

For instance, a new shop with new umu-ahja (ages of 3 and 8 yrs.), some constraints of this age group include (a) Nostalgic energy in early months, inherent childishness (comprises of untargetted and undirected energies): chaotic sleeping patterns, food pattern, and desires for the first few months persist. But discipleship (hands-on, disciplinary

actions, corrections, task handling, rewards) between 1 and 5 years changes their energies to focused excitement; next is integrity and trust test; Each nwa-ahja has a position and cost affected by such constraints and uncertainty in capacity, inductiveness, and reluctance. The population is generated using (3).

$$pop = \Phi(M, D) \quad (3)$$

The  $\Phi$  is a random generator, M is the market population (pop), and D is the dimension. The pseudocode is shown below.

#### Initial parameters

Initialize the structure for the empty individuals

Initialize population array

while (not termination) Do

    generate uniform random population with

        bounded size of market

    evaluate the cost of the individual

    update individual population

end while

return best solution

#### E. Mathematics of Igba-oso-ahja

Fig. 5(a) in 3-dimensional space during the search is shown in 2-Dimension as Fig. 5 (b) as agents move to satisfy customers' scarce demand for "gold" and "leather" as in Fig. 4 to exploit S1 and explore S2.

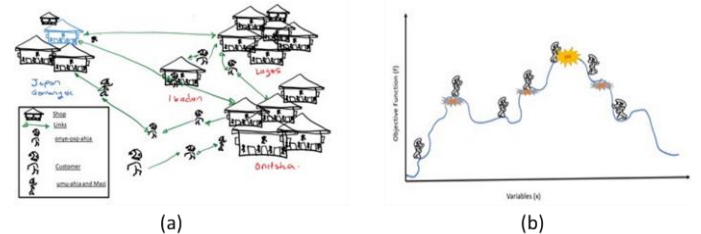


Fig. 5. Igba-oso-ahja cooperation by to find gold.

They cooperate, meet set thresholds, and satisfy the customer to get his Gold. Initializing a new Mazi, a new shop, and his contribution to the solution space will be given by (4).

$$U_n = f(X_1) \quad (2)$$

Combining equation (1) and (2);

$$U_n \leftarrow \begin{cases} x_i \in \{x_r\} & \text{mazi with new umu-ahja} \\ x_i \in X_1 & \text{1mazi, without umu-ahja} \end{cases} \quad (3)$$

Some of the major constraints (3) as mentioned in Section II. n; the number of generations – the stopping criteria, r indicates x is random. Mazi cost alone in trade without umu-ahja in the cycle of Ahja days (6) resolves to a fitness vector:

$$U_n = X_{n-1} * (P(i,:)) \quad (6)$$

Where  $P(i,:)$  is the cost of Mazi in the population of P,  $i$  cycles of Ahja days, and the transmute of Mazi's energies via the training processes. At the same time, the  $\mu\mu$ -ahja adopt the emitted energies originating from multidiscipline like phycology, social tactics, resilience, experience, transactional techniques, relationship with customers, banks, etc., The differences compared to theirs cut across domains and, by analogy, involve transduction [28]. This process is given by a resemblance of balancing potentials and kinetics.

$$T = rand(1, D) * (P(p,:)) - \frac{1}{2} * Ef * (X_r) \quad (4)$$

Where  $P(p,:)$  in (4) is the cost of the new population at time p,  $Ef$  is the energy factor, while the  $X_r$  random  $\mu\mu$ -ahja cognitive state of five analogous Bayesian energies interacting actively in a shop.

$$X_r = X_{r+i} + 1 / 8 * Ef * (X_{r+i-1}) \quad (5)$$

$$X_{r+i-1} = X_{r+\Delta(i-1)} + 1 / 40 * Ef * X_{r+\Delta(i-1)} \quad (6)$$

$$X_{r+\Delta(i-1)} = 1 / 80 * X_{r+\Delta(i-n)} \quad (7)$$

Equations (5), (6) and (7) are all nonlinear cognitive vectors, and ratios of series [1/8, 1/40, and 1/80] of time divisions (could take any ratio as they are probabilities of random events, recall a state ranges from 0 to 1),  $i, r$  remain the same; visually, a huge network ensues as shown below.

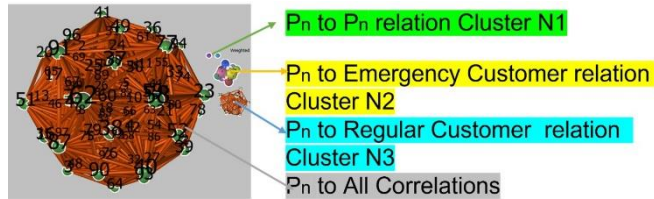


Fig. 6. Cognitive correlations between nwa-ahja (PN) to any.

Fig. 6 clusters N1  $\mu\mu$ -ahja with each other, and N2 and N3 are the clusters with customers. The probabilistic behavior of the interaction (edges) shows the very interesting transition of the  $\mu\mu$ -ahja (nodes) in the network in Fig. 7.

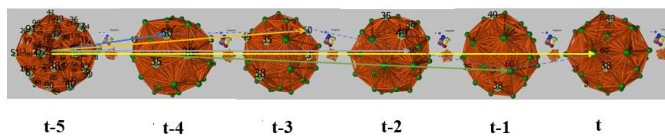


Fig. 7. Progress of cognitive signature on  $\mu\mu$ -ahja character.

Current time  $t$ ; the previous timestamps as the selected node (60). Search in a generation gives (11).

$$X_r = X_{r+i} + 1 / 8 * Ef * (X_{r+\Delta(i-1)} + \dots + 1 / 40 * Ef (1 / 80 * X_{r+\Delta(i-n)})) \quad (8)$$

Where updates at  $r+i$  taken during iteration. The compact dynamics (12); mimics a rhythmic nodding to music and stratagem - igba, which gives.

$$U_n = U_{n-1} + T_{n-1} * X_r \quad (12)$$

Where  $U_n$  is a vector of emergent solutions. The threshold facilitates  $\mu\mu$ -ahja exploration; disciple-Rhythm -  $rD$  known as discipleship compliance, given by (13).

$$U(ij) = \begin{cases} u_m^j & \text{if } m \leq rD \text{ OR } j = \delta \\ x_m^j & \text{if } m > rD \text{ AND } j \neq \delta \end{cases} \quad (9)$$

Where  $m$  is a random number [0,1], delta  $\delta$  has the same dimension and size as the solution but is pseudorandom. This cooperation serves as the bond linking one source to another [23, 29]. Mazi; sometimes sacrifices profit for an improved - customer base and to escape the local optima trap by analogy as Igba-oso-ahja updates; objectives of the fitness bound the strategy as given in Fig. 8.

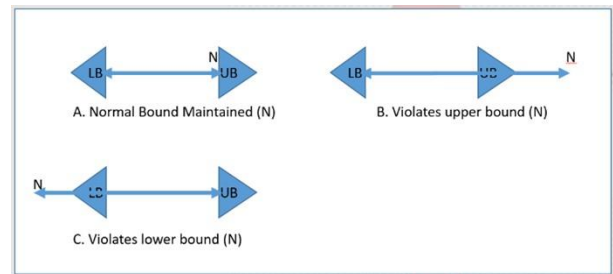


Fig. 8. Boundary strategy.

Constraints are bounded as in A, N collapse to upper-bound (UB) in B, and N collapse to lower-bound (LB) as shown in Fig. 8. Finally, methodic update results to best solutions shown.

$$X_i = \begin{cases} U_i \\ f_i = f_{U_i} \end{cases} \text{ if } f_{U_i} < f_i \quad (10)$$

$X$  and  $f$  remains the same if  $f_{U_i} > f_i$

Where (10);  $f_{U_i}$  is the fitness function from the best cost of the discipleship and adjustments made (error correction), while  $f_i$  is the best solution fitness of the original objective function, which is optimal.

#### F. Graphical Flow of QMOA

The  $\mu\mu$ -ahja can be considered as moving particles [30-32]. Mazi realization comes after generations of successful cycles [33]; rather than unhealthy competition, all  $\mu\mu$ -ahja depends on each other; The main body's pseudocode (2) during iteration is as follows:

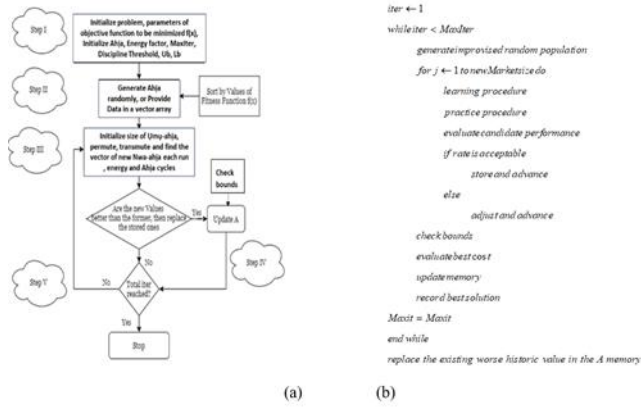


Fig. 9. Qdigbo metaheuristic optimization algorithm - QMOA and pseudocode.

The flow chart of Fig. 9 shows the methodology for applying the QMOA algorithm. The subsequent sections discuss applications.

### III. APPLICATION OF QMOA ON BENCHMARK FUNCTIONS

Most metaheuristic algorithms use Pattern Matrix, and the solutions are identified as those that improved through the number of generations up until the convergence time of the simulation. QMOA inherent energy synergy principles.

- 1) Default parameters are used
- 2) 30 independent runs were used for Unconstrained Benchmark, 50 for constraint functions. The parameters for the engineering designs are as stated in the referenced literature provided
- 3) The total number of cost function evaluations is  $1000 \cdot n \cdot M$ , where  $M = 10$  is the number of iterations. The logarithmic Scale was considered for visualization due to its convenience and compact.
- 4) For the Constraint problems as depicted by the competition, a solution value less than  $10^{-8}$  is treated as zero; several performance indicators for solution values are used: best, worst, mean, and standard deviation (Std). Test for convergence time also provided;

#### A. Experiments and Comparison of Results

QMOA is compared with five of the best similar algorithms as shown in Table II. Their codes are in the open domain/available online. The choice of only five is being mindful also of the limited space constraints to publish results.

TABLE II. ALGORITHMS USED FOR VALIDATION

S/N	Algorithm	Ref	Category
1	Harris Hawks Optimization (HHO)-- 2019	[34]	Novel Idea
2	Moth Search (MS)-- 2018	[35]	Novel Idea
3	Elephant Herd Optimization (EHO)-- 2015	[36]	Novel Idea
4	LSHADE-SPACMA ( A2) - 2017	[37]	Hybrid/Modified
5	EBOwithCMAR ( A3) ---2017	[38]	Hybrid/Modified

The list in Table II is a competitive group; notably, 4 – 5 won the CEC 2017 competition [39, 40].

#### B. Experiment 1: Difficult Unconstrained Benchmark Functions

QMOA is validated on the existing established algorithms listed in Table II with about 30 difficult functions chosen with modality (unimodal to check and confirm exploitation strength, multimodal for diversity or exploratory capability of QMOA), Separability (possible separable and non-separable) and then multi-dimensionality (confirming search and exploratory strength of QMOA). The performance averages are visualized using boxplots. Further, the significance and statistical students test (t-test) was conducted for all algorithms, with a time complexity test. A subset of test benchmark functions with varying degrees of difficulty is used to substantiate that QMOA can exploit and explore the solution space and find the solutions for optimum. In Table III, unconstrained benchmark test functions are categorized in modality, Separability, and Dimensionality ( $N$ ), also:  $M$  is the modality,  $\theta$  – Uni-modal;  $I$  – Multimodal,  $S$  is the Separability,  $\theta$  – Non-Separable;  $I$  – Separable.

TABLE III. FUNCTIONS, GLOBAL OPTIMAL VALUES, BOUNDS, AND DIMENSIONS

Fun(fn)	Fun-name	SD	F(x*)	N	M	S
F1	Step	[-5.12, 5.12] <sup>n</sup>	0	30	0	1
F2	Sphere	[-1, 1] <sup>n</sup>	0	30	0	1
F3	Sum Square	[-5.12, 5.12] <sup>n</sup>	0	30	0	1
F4	Quartic	[-6.0, 6.0] <sup>n</sup>	0	30	0	1
F5	Beale	[-5.0, 5.0] <sup>n</sup>	0	2	1	0
F6	Easom	[-100.0, 100.0] <sup>n</sup>	-1	2	1	0
F7	Matyas	[-10.0, 10.0] <sup>n</sup>	0	2	1	0
F8	Colville	[-10.0, 10.0] <sup>n</sup>	0	4	1	0
F9	Zakharov	[-5.0, 5.0] <sup>n</sup>	0	30	1	0
F10	Schwefel 2.2	[-10.0, 10.0] <sup>n</sup>	0	30	1	0
F11	Schwefel 1.2	[-10.0, 10.0] <sup>n</sup>	0	30	1	0
F12	Dixon Price	[-10.0, 10.0] <sup>n</sup>	0	30	1	0
F13	Bohachevsky 1	[-100.0, 100.0] <sup>n</sup>	0	2	1	1
F14	Booth	[-10, 10] <sup>n</sup>	0	2	1	1
F15	Holder Table	[-10 10] <sup>n</sup>	-19.2085	2	1	1
F16	Michalewicz 2	[0.0, π] <sup>n</sup>	-1.8013	2	1	1
F17	Michalewicz 5	[0.0, π] <sup>n</sup>	-4.6877	5	1	1
F18	Michalewicz 10	[0.0, π] <sup>n</sup>	-9.6602	10	1	1
F19	Rastrigin	[-5.12, 5.12] <sup>n</sup>	0	5	1	1
F20	Schaffer2	[-100, 100] <sup>n</sup>	0	2	1	0
F21	Schaffer 4	[-100.0, 100.0] <sup>n</sup>	0	4	1	0
F22	Schaffer 6	[-100.0, 100.0] <sup>n</sup>	0	6	1	0
F23	SixHumpCamelBack	[-5,5]	-1.0316	2	1	1
F24	Bohachevsky 2	[-100.0, 100.0] <sup>n</sup>	0	2	1	0
F25	Bohachevsky 3	[-100.0, 100.0] <sup>n</sup>	0	3	1	1
F26	Shubert	[-10.0, 10.0] <sup>n</sup>	-186.73	5	1	1
F27	Drop Wave	[-5.12, 5.12] <sup>n</sup>	-1	2	0	1
F28	Rosenbrock	[-6.0, 6.0] <sup>n</sup>	0	2	0	0
F29	Griewank	[-600.0, 600.0] <sup>n</sup>	0	30	1	0
F30	Ackley	[-32.0, 32.0] <sup>n</sup>	0	2	1	0

C. Benchmark – Unimodal and Separable Functions

To tighten the competitiveness, we identified the algorithms with the highest performances with a t-value above 0.05. QMOA and A3 (all best solutions in BOLD font) lead with equal best performance as shown in Table IV. The MS was next, followed by A1, HHO, and WFS.

QMOA and A3 leading showed good exploration, and exploitation strength, particularly of QMOA obtained the best optimal objective solutions before the completion of generations. Fig. 10 also shows consistent distributions with fewer outliers.

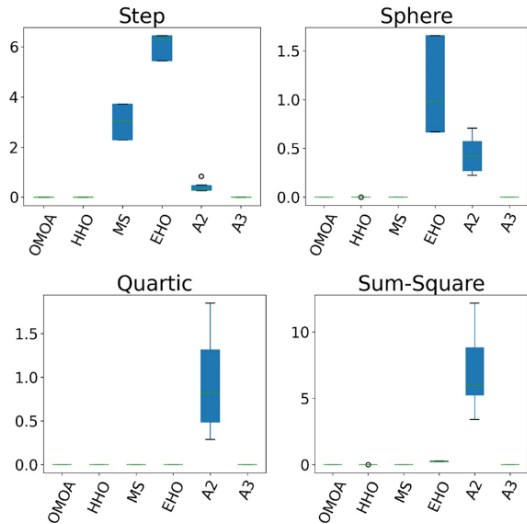


Fig. 10. Boxplot for unimodal and separable functions.

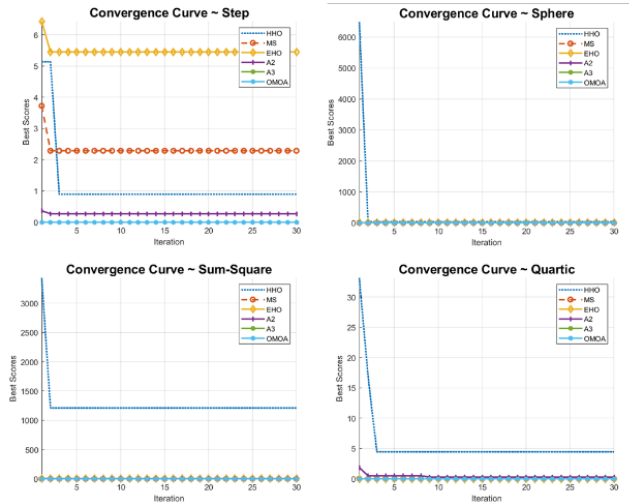


Fig. 11. Comparison of convergence curve for unimodal and separable functions.

Fig. 11 convergence comparison shows that QMOA, A3, A2, and WFS have faster and best convergences to the optimum in these problems while HHO lagged behind most of the time.

D. Unimodal and Non-Separable

The functions in this category include 30-dimensional problems Zakharov to Dixon Price with great complexity. Table V shows QMOA and A3 tops, followed by A2, A1, HHO, MS, and WFS. Besides Beale, which did not yield a better result, QMOA got even Easom, a problem with inherent complex nature.

TABLE IV. UNIMODAL AND SEPARABLE FUNCTIONS RESULTS

F(n)	Measure	QMOA	HHO	MS	EHO	A2	A3
F1	Best	<b>0.0E+00</b>	1.3E-06	2.3E+00	5.5E+00	1.5E-01	<b>0.0E+00</b>
n = 30	Worst	<b>0.0E+00</b>	7.6E-05	3.7E+00	6.4E+00	8.5E-01	<b>0.0E+00</b>
f = 0	Mean	<b>0.0E+00</b>	2.2E-05	3.0E+00	6.1E+00	4.1E-01	<b>0.0E+00</b>
	Sd	<b>0.0E+00</b>	3.0E-05	3.1E+00	6.1E+00	4.3E-01	<b>0.0E+00</b>
F2	Best	<b>0.0E+00</b>	0.0E+00	<b>0.0E+00</b>	6.7E-01	1.6E-01	<b>0.0E+00</b>
n = 30	Worst	<b>0.0E+00</b>	1.1E-07	<b>0.0E+00</b>	1.7E+00	7.1E-01	<b>0.0E+00</b>
f = 0	Mean	<b>0.0E+00</b>	1.1E-08	<b>0.0E+00</b>	1.1E+00	4.2E-01	<b>0.0E+00</b>
	Sd	<b>0.0E+00</b>	2.7E-08	<b>0.0E+00</b>	1.2E+00	4.1E-01	<b>0.0E+00</b>
F3	Best	<b>0.0E+00</b>	<b>0.0E+00</b>	0.0E+00	2.2E-01	2.5 E+00	<b>0.0E+00</b>
n = 30	Worst	<b>0.0E+00</b>	<b>1.2E-07</b>	7.2E-08	2.9E-01	1.2E+01	<b>0.0E+00</b>
f = 0	Mean	<b>0.0E+00</b>	<b>6.4E-09</b>	4.3E-08	2.5E-01	6.3E+00	<b>0.0E+00</b>
	Sd	<b>0.0E+00</b>	<b>2.4E-08</b>	5.3E-08	2.5E-01	6.7E+00	<b>0.0E+00</b>
F4	Best	<b>0.0E+00</b>	<b>0.0E+00</b>	<b>0.0E+00</b>	1.2E-06	2.1E-01	<b>0.0E+00</b>
n = 30	Worst	<b>0.0E+00</b>	<b>0.0E+00</b>	<b>0.0E+00</b>	2.2E-06	5.0E+00	<b>0.0E+00</b>
f = 0	Mean	<b>0.0E+00</b>	<b>0.0E+00</b>	<b>0.0E+00</b>	1.6E-06	1.3E+00	<b>0.0E+00</b>
	Sd	<b>0.0E+00</b>	<b>0.0E+00</b>	<b>0.0E+00</b>	1.6E-06	1.5E+00	<b>0.0E+00</b>

TABLE V. UNIMODAL AND NON-SEPARABLE RESULTS FOR TESTED ALGORITHMS

Function	Measure	QMOA	HHO	MS	EHO	A2	A3
<b>F5</b>	Best	5.7E-07	<b>0.0E+00</b>	6.0E-05	1.1E-05	0.0E+00	<b>0.0E+00</b>
<b>n = 2</b>	Worst	1.7E-01	<b>0.0E+00</b>	7.2E-04	2.3E-02	1.5E-05	<b>0.0E+00</b>
<b>f = 0</b>	Mean	4.7E-02	<b>0.0E+00</b>	2.9E-04	7.9E-03	1.0E-06	<b>0.0E+00</b>
	Sd	4.8E-02	<b>0.0E+00</b>	4.2E-04	1.3E-02	2.9E-06	<b>0.0E+00</b>
<b>F6</b>	Best	<b>-9.6E-01</b>	0.0E+00	7.7E-04	1.2E-01	0.0E+00	0.0E+00
<b>n = 2</b>	Worst	<b>-1.7E-235</b>	1.9E-05	1.3E-03	5.7E-01	0.0E+00	0.0E+00
<b>f = -1</b>	Mean	<b>-3.1E-01</b>	1.1E-06	1.0E-03	3.8E-01	0.0E+00	0.0E+00
	Sd	<b>3.6E-01</b>	3.6E-06	1.1E-03	4.2E-01	0.0E+00	0.0E+00
<b>F7</b>	Best	<b>0.0E+00</b>	<b>0.0E+00</b>	<b>0.0E+00</b>	4.3E-07	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>n = 2</b>	Worst	<b>0.0E+00</b>	<b>0.0E+00</b>	<b>0.0E+00</b>	3.0E-06	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>f = 0</b>	Mean	<b>0.0E+00</b>	<b>0.0E+00</b>	<b>0.0E+00</b>	1.9E-06	<b>0.0E+00</b>	<b>0.0E+00</b>
	Sd	<b>0.0E+00</b>	<b>0.0E+00</b>	<b>0.0E+00</b>	2.2E-06	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>F8</b>	Best	<b>0.0E+00</b>	1.6E-06	1.8E-01	1.1E+00	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>n = 4</b>	Worst	<b>0.0E+00</b>	7.7E-01	2.9E-01	3.2E+00	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>f = 0</b>	Mean	<b>0.0E+00</b>	3.4E-02	2.4E-01	1.9E+00	<b>0.0E+00</b>	<b>0.0E+00</b>
	Sd	<b>0.0E+00</b>	1.4E-01	2.4E-01	2.1E+00	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>F9</b>	Best	<b>0.0E+00</b>	0.0E+00	<b>0.0E+00</b>	3.6E-02	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>n = 30</b>	Worst	<b>0.0E+00</b>	7.7E-07	<b>1.3E-08</b>	9.1E-02	5.1E-08	<b>0.0E+00</b>
<b>f = 0</b>	Mean	<b>0.0E+00</b>	8.0E-08	<b>4.3E-09</b>	6.0E-02	1.4E-08	<b>0.0E+00</b>
	Sd	<b>0.0E+00</b>	1.8E-07	<b>7.4E-09</b>	6.4E-02	2.1E-08	<b>0.0E+00</b>
<b>F10</b>	Best	<b>0.0E+00</b>	4.3E-07	<b>3.4E-05</b>	4.8E-01	6.4E-03	8.7E-06
<b>n = 30</b>	Worst	<b>0.0E+00</b>	2.7E-04	4.6E-04	6.2E-01	2.7E-02	4.3E-05
<b>f = 0</b>	Mean	<b>0.0E+00</b>	4.7E-05	1.9E-04	5.4E-01	1.8E-02	1.8E-05
	Sd	<b>0.0E+00</b>	7.6E-05	2.7E-04	5.5E-01	1.9E-02	2.1E-05
<b>F11</b>	Best	<b>0.0E+00</b>	0.0E+00	8.0E-08	3.4E+00	8.2E-02	<b>0.0E+00</b>
<b>n = 30</b>	Worst	<b>0.0E+00</b>	1.9E-03	3.8E-07	4.2E+00	5.5E-01	<b>0.0E+00</b>
<b>f = 0</b>	Mean	<b>0.0E+00</b>	9.4E-05	2.7E-07	3.8E+00	3.0E-01	<b>0.0E+00</b>
	Sd	<b>0.0E+00</b>	3.6E-04	3.0E-07	3.8E+00	3.4E-01	<b>0.0E+00</b>
<b>F12</b>	Best	<b>3.1E-09</b>	8.6E-02	6.7E-01	1.1E+00	6.7E-01	6.7E-01
<b>n = 30</b>	Worst	<b>6.7E-04</b>	2.6E-01	7.0E-01	1.3E+00	6.7E-01	6.7E-01
<b>f = 0</b>	Mean	<b>3.0E-05</b>	2.4E-01	6.9E-01	1.2E+00	6.7E-01	6.7E-01
	Sd	<b>1.2E-04</b>	2.4E-01	6.9E-01	1.2E+00	6.7E-01	6.7E-01



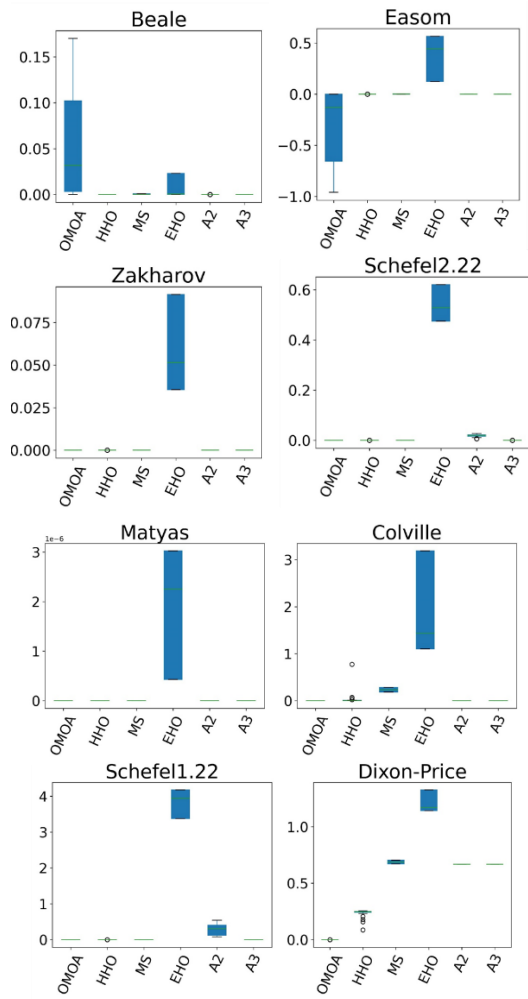


Fig. 12. Boxplot of unimodal and non-separable results.

The boxplot of Fig. 12 shows that the mean solutions distribution of the data of QMOA and A3 are tight, with little outliers equalling minimal deviation.

The convergence comparison of Fig. 13 confirms the summary made at the beginning of the subsection.

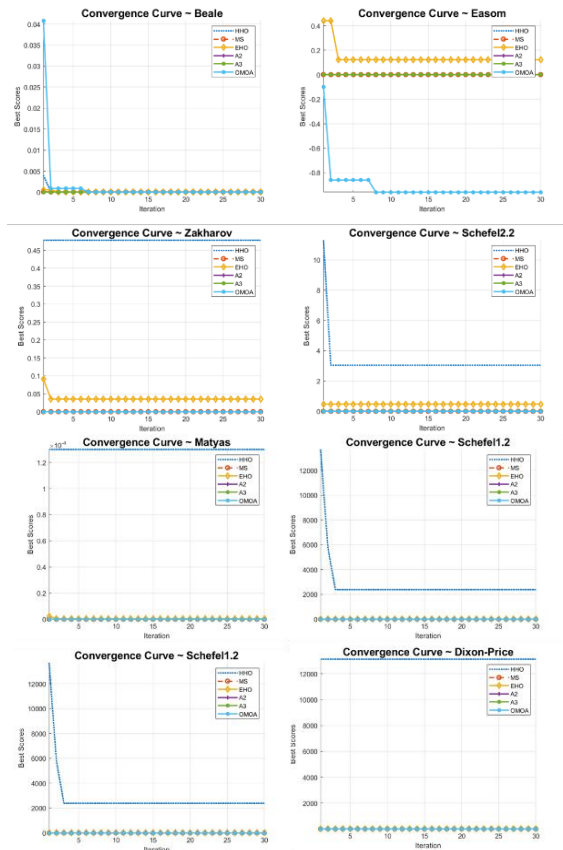


Fig. 13. Comparison of convergence curves for unimodal and non-separable function.

### E. Multimodal and Separable

Complex structures, multiple, unequal hilltops, and valleys-shaped functions are tested as shown in Table VI. Besides the booth function, QMOA had remarkable exploratory abilities for the dimensionalities above  $n = 2$  (i.e.,  $n = 5, 10, 30$ ) of the last three functions while tracking deeper than values provided by the global optima in literature for Holder Table, Michalewicz (2, 5, and 10). Rastrigin ( $n = 30$ ) was also explored optimally by QMOA and HHO.

TABLE VI. MULTIMODAL AND SEPARABLE RESULTS

Function	Measure	QMOA	HHO	MS	EHO	A2	A3
F13	Best	0.0E+00	0.0E+00	2.8E-08	5.4E-03	0.0E+00	0.0E+00
n = 2	Worst	0.0E+00	2.7E-08	2.5E-07	1.5E-02	0.0E+00	0.0E+00
f = 0	Mean	0.0E+00	3.3E-09	1.7E-07	1.0E-02	0.0E+00	0.0E+00
	Sd	0.0E+00	8.5E-09	2.0E-07	1.1E-02	0.0E+00	0.0E+00
F14	Best	1.1E-04	0.0E+00	8.7E-06	3.8E-05	0.0E+00	0.0E+00
n = 2	Worst	2.6E-02	4.1E-04	4.1E-04	3.7E-02	0.0E+00	0.0E+00
f = 0	Mean	5.5E-03	9.1E-05	2.4E-04	2.0E-02	0.0E+00	0.0E+00
	Sd	5.9E-03	1.5E-04	3.0E-04	2.6E-02	0.0E+00	0.0E+00
F15	Best	-5.0E+04	0.0E+00	-1.8E+01	-1.7E+01	5.3E-06	0.0E+00
n = 2	Worst	-2.5E+04	1.3E-05	-1.1E+01	-1.1E+01	3.4E-04	0.0E+00
f = -19.2085	Mean	-3.9E+04	1.4E-06	-1.5E+01	-1.4E+01	1.0E-04	0.0E+00

	Sd	<b>1.2E+04</b>	3.4E-06	<b>2.8E+00</b>	<b>2.7E+00</b>	1.1E-04	0.0E+00
<b>F16</b>	Best	<b>-1.8E+00</b>	0.0E+00	1.2E-03	6.5E-03	0.0E+00	0.0E+00
<b>n = 2</b>	Worst	<b>-1.7E+00</b>	2.3E-06	7.4E-03	3.0E-02	0.0E+00	0.0E+00
<b>f = -1.8013</b>	Mean	<b>-1.8E+00</b>	2.9E-07	3.3E-03	1.7E-02	0.0E+00	0.0E+00
	Sd	<b>1.4E-02</b>	5.5E-07	4.4E-03	2.0E-02	0.0E+00	0.0E+00
<b>F17</b>	Best	<b>-4.0E+00</b>	1.5E-02	2.4E-01	9.0E-01	2.6E-06	0.0E+00
<b>n = 5</b>	Worst	<b>-2.8E+00</b>	1.2E+00	1.3E+00	1.2E+00	1.4E-04	4.4E-02
<b>f = -4.6877</b>	Mean	<b>-3.6E+00</b>	4.7E-01	8.5E-01	1.1E+00	4.0E-05	8.7E-03
	Sd	<b>2.8E-01</b>	6.1E-01	9.6E-01	1.1E+00	6.1E-05	1.8E-02
<b>F18</b>	Best	<b>-5.5E+00</b>	1.6E+00	3.4E+00	3.4E+00	8.6E-01	6.5E-01
<b>n = 10</b>	Worst	<b>-4.2E+00</b>	3.8E+00	4.9E+00	4.2E+00	1.3E+00	1.4E+00
<b>f = -9.6602</b>	Mean	<b>-4.7E+00</b>	2.9E+00	4.2E+00	3.9E+00	1.1E+00	1.1E+00
	Sd	<b>3.5E-01</b>	2.9E+00	4.2E+00	4.0E+00	1.1E+00	1.1E+00
<b>F19</b>	Best	<b>0.0E+00</b>	0.0E+00	0.0E+00	8.3E-01	9.2E+01	1.5E+01
<b>n = 30</b>	Worst	<b>0.0E+00</b>	3.3E-07	2.2E-07	1.2E+00	1.1E+02	7.3E+01
<b>f = 0</b>	Mean	<b>0.0E+00</b>	2.1E-08	1.2E-07	1.0E+00	1.0E+02	3.5E+01
	Sd	<b>0.0E+00</b>	6.5E-08	1.5E-07	1.0E+00	1.0E+02	3.9E+01

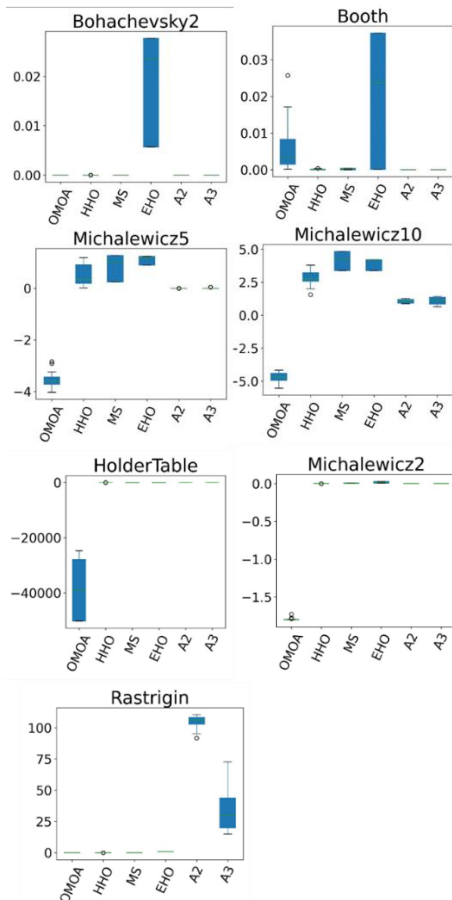


Fig. 14. Multimodal and separable boxplot.

The visuals of the boxplot in Fig. 14 show mean solutions of QMOA adequately located in the region with very few deviations and outliers.

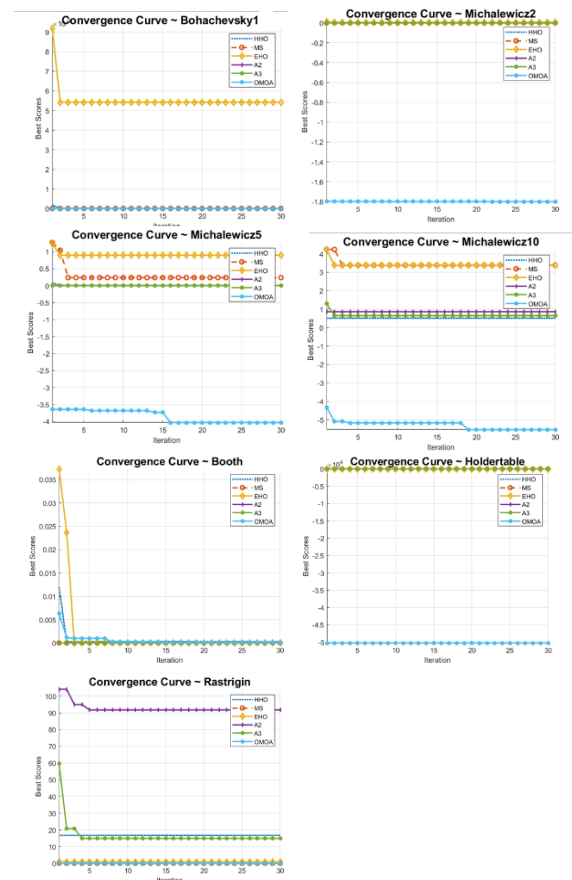


Fig. 15. Convergence curves for multimodal and separable function.

In Fig. 15, QMOA had made extra-advance to explore for solutions far better than all the compared algorithms in these problems. Even some of the solutions were far better optimum that set global values as the Holder Table model.

F. Multimodal and Non-Separable

Table VII shows QMOA led the exploration alongside A3, A2, and HHO though the depth of the troughs in Six-Hump

Camel, Shubert, and Drop Wave seems to have shown that QMOA dived deeper than the others with the peaks of Schaffer 4 also.

TABLE VII. MULTIMODAL AND NON-SEPARABLE RESULTS

Function	Measure	QMOA	HHO	MS	EHO	A2	A3
<b>F20</b>	Best	<b>0.0E+00</b>	<b>0.0E+00</b>	<b>0.0E+00</b>	3.3E-07	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>n = 2</b>	Worst	<b>4.3E-01</b>	<b>0.0E+00</b>	<b>0.0E+00</b>	6.6E-07	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>f = 0</b>	Mean	<b>9.3E-02</b>	<b>0.0E+00</b>	<b>0.0E+00</b>	4.9E-07	<b>0.0E+00</b>	<b>0.0E+00</b>
	Sd	<b>1.2E-01</b>	<b>0.0E+00</b>	<b>0.0E+00</b>	5.1E-07	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>F21</b>	Best	<b>0.0E+00</b>	<b>0.0E+00</b>	<b>5.8E-08</b>	1.3E-06	<b>1.0E-08</b>	<b>0.0E+00</b>
<b>n = 4</b>	Worst	<b>1.6E-02</b>	<b>1.1E-05</b>	<b>7.4E-06</b>	1.5E-04	<b>2.5E-05</b>	<b>0.0E+00</b>
<b>f = 0.29259</b>	Mean	<b>1.3E-03</b>	<b>1.2E-06</b>	<b>2.6E-06</b>	5.3E-05	<b>6.1E-06</b>	<b>0.0E+00</b>
	Sd	<b>3.7E-03</b>	<b>3.2E-06</b>	<b>4.3E-06</b>	8.7E-05	<b>9.3E-06</b>	<b>0.0E+00</b>
<b>F22</b>	Best	9.7E-03	0.0E+00	0.0E+00	5.2E+00	6.7E+00	<b>5.5E+00</b>
<b>n = 6</b>	Worst	4.0E-01	2.3E-03	2.2E-08	6.1E+00	9.1E+00	<b>7.5E+00</b>
<b>f = 0</b>	Mean	1.7E-01	9.0E-05	7.3E-09	5.7E+00	8.1E+00	<b>6.6E+00</b>
	Sd	1.2E-01	4.3E-04	1.3E-08	5.8E+00	8.1E+00	<b>6.7E+00</b>
<b>F23</b>	Best	<b>-1.0E+00</b>	0.0E+00	8.6E-07	1.7E-04	0.0E+00	0.0E+00
<b>n = 2</b>	Worst	<b>-1.0E+00</b>	1.7E-08	3.4E-05	1.9E-03	1.1E-07	0.0E+00
<b>f = -1.03163</b>	Mean	<b>-1.0E+00</b>	2.2E-09	1.6E-05	8.2E-04	3.2E-08	0.0E+00
	Sd	<b>9.6E-05</b>	5.5E-09	2.1E-05	1.1E-03	5.3E-08	0.0E+00
<b>F24</b>	Best	<b>0.0E+00</b>	<b>0.0E+00</b>	1.2E-08	5.7E-03	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>n = 2</b>	Worst	<b>0.0E+00</b>	<b>3.3E-07</b>	2.1E-07	2.8E-02	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>f = 0</b>	Mean	<b>0.0E+00</b>	<b>2.5E-08</b>	1.2E-07	1.9E-02	<b>0.0E+00</b>	<b>0.0E+00</b>
	Sd	<b>0.0E+00</b>	<b>7.7E-08</b>	1.4E-07	2.1E-02	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>F25</b>	Best	<b>0.0E+00</b>	0.0E+00	0.0E+00	3.0E-03	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>n = 2</b>	Worst	<b>0.0E+00</b>	9.0E-06	1.4E-08	9.5E-03	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>f = 0</b>	Mean	<b>0.0E+00</b>	6.2E-07	4.6E-09	7.2E-03	<b>0.0E+00</b>	<b>0.0E+00</b>
	Sd	<b>0.0E+00</b>	1.9E-06	8.0E-09	7.8E-03	<b>0.0E+00</b>	<b>0.0E+00</b>
<b>F26</b>	Best	<b>-1.9E+02</b>	<b>-9.0E-06</b>	1.1E-02	2.3E-01	<b>-6.0E-06</b>	<b>-9.0E-06</b>
<b>n = 2</b>	Worst	<b>-1.9E+02</b>	2.7E-04	3.2E-02	1.8E+00	<b>8.2E-02</b>	<b>-9.0E-06</b>
<b>f = -186.73</b>	Mean	<b>-1.9E+02</b>	2.4E-05	2.2E-02	1.2E+00	<b>2.8E-02</b>	<b>-9.0E-06</b>
	Sd	<b>9.3E-02</b>	6.5E-05	2.3E-02	1.4E+00	<b>4.0E-02</b>	<b>0.0E+00</b>
<b>F27</b>	Best	<b>-1.0E+00</b>	0.0E+00	0.0E+00	3.9E-06	0.0E+00	0.0E+00
<b>n = 2</b>	Worst	<b>-1.0E+00</b>	0.0E+00	1.8E-07	3.8E-05	0.0E+00	0.0E+00
<b>f = -1</b>	Mean	<b>-1.0E+00</b>	0.0E+00	6.6E-08	2.6E-05	0.0E+00	0.0E+00
	Sd	<b>0.0E+00</b>	0.0E+00	1.1E-07	3.0E-05	0.0E+00	0.0E+00
<b>F28</b>	Best	2.9E+01	1.1E-03	2.9E+01	3.8E+01	2.7E+01	<b>0.0E+00</b>
<b>n = 30</b>	Worst	2.9E+01	2.5E-01	2.9E+01	4.0E+01	2.8E+01	4.0E+00
<b>f = 0</b>	Mean	2.9E+01	4.5E-02	2.9E+01	3.9E+01	2.8E+01	1.2E+00
	Sd	9.0E-02	7.1E-02	2.9E+01	3.9E+01	2.8E+01	2.2E+00
<b>F29</b>	Best	<b>0.0E+00</b>	0.0E+00	0.0E+00	9.3E-01	6.1E-04	<b>0.0E+00</b>
<b>n = 30</b>	Worst	<b>0.0E+00</b>	1.1E-06	2.2E-08	1.0E+00	2.2E-03	<b>0.0E+00</b>
<b>f = 0</b>	Mean	<b>0.0E+00</b>	6.9E-08	7.3E-09	9.7E-01	1.3E-03	<b>0.0E+00</b>
	Sd	<b>0.0E+00</b>	2.1E-07	1.3E-08	9.8E-01	1.4E-03	<b>0.0E+00</b>
<b>F30</b>	Best	<b>8.9E-16</b>	1.1E-06	1.7E-05	4.3E-01	3.9E-03	1.9E-06
<b>n = 30</b>	Worst	<b>8.9E-16</b>	1.0E-04	3.6E-05	5.4E-01	6.2E-03	1.7E+00
<b>f = 0</b>	Mean	<b>8.9E-16</b>	1.8E-05	2.4E-05	4.8E-01	5.0E-03	7.2E-01
	Sd	<b>8.9E-16</b>	2.7E-05	2.5E-05	4.9E-01	5.0E-03	9.0E-01

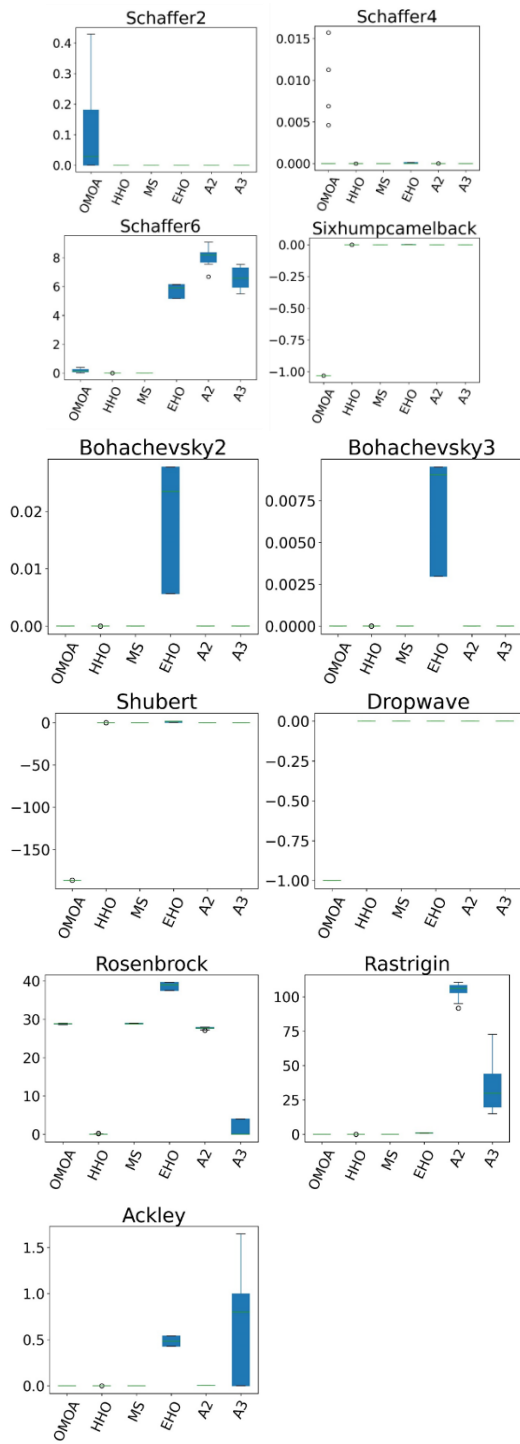


Fig. 16. Multimodal and non-separable boxplot.

The boxplot in Fig. 16 shows the clear visuals, confirms OMOA better. The exploratory ability of the OMOA is evident from mean solutions distribution and standard deviations.

The convergence is a reflection of the ability of the tree depth of the network of markets embedded in the model visualized in Fig. 17.

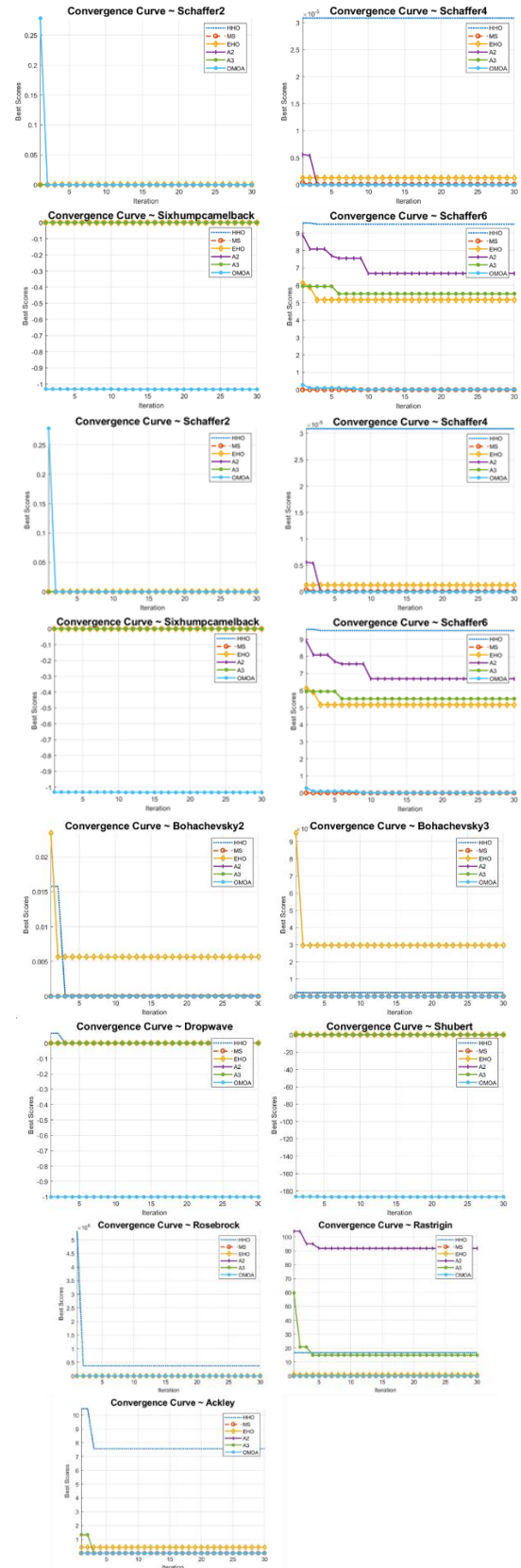


Fig. 17. Convergence curves for multimodal and non-separable function.

G. Statistical Test and Significance

Table VIII presents the entire statistical hypothesis test carried out to confirm the difference in mean and significance validation in the distribution of solutions by the algorithms on the 30 unconstraint benchmark functions.

Table IX is the summary of the test conducted to prove the hypothesis of the performances of the experiment; **1**: means QMOA (in black ink) is more significant, **-1**: gives better significance to the contender (another algorithm), while **0**: depicts no significant difference in performance (contender, equal, QMOA).

TABLE VIII. T-TEST FOR PERFORMANCE AND HYPOTHESIS

S/ N	Algorithm/ Function	HHO			MS			EHO			A2			A3		
		t-value	p-value	Sig <sub>n</sub>	t-value	p-value	Sig <sub>n</sub>	t-value	p-value	Sig <sub>n</sub>	t-value	p-value	Sig <sub>n</sub>	t-value	p-value	Sig <sub>n</sub>
1	Step	-5.8E+00	2.5E-06	1	-2.8E+01	2.1E-22	1	-7.5E+01	8.7E-35	1	-1.3E+01	8.9E-14	1	nan	nan	0
2	Sphere	-2.3E+00	3.1E-02	1	nan	nan	0	-1.4E+01	9.6E-15	1	-1.5E+01	5.4E-15	1	nan	nan	0
3	Sum-Square	-1.5E+00	1.4E-01	1	-7.5E+00	2.8E-08	1	-4.5E+01	2.4E-28	1	-1.4E+01	1.0E-14	1	nan	nan	0
4	Quartic	nan	nan	0	nan	nan	0	-2.0E+01	1.9E-18	1	-9.7E+00	1.4E-10	1	nan	nan	0
5	Beale	5.3E+00	9.9E-06	-1	5.3E+00	1.1E-05	-1	4.1E+00	3.3E-04	-1	5.3E+00	9.9E-06	1	5.3E+00	9.9E-06	-1
6	Easom	-4.6E+00	7.5E-05	1	-4.6E+00	7.2E-05	1	-8.0E+00	9.1E-09	1	-4.6E+00	7.5E-05	1	-4.6E+00	7.5E-05	1
7	Matyas	nan	nan	0	nan	nan	0	-9.4E+00	2.6E-10	1	nan	nan	0	nan	nan	0
8	Colville	-5.3E+00	1.1E-05	1	-5.1E+00	2.0E-05	1	-3.0E+01	1.7E-23	1	nan	nan	0	nan	nan	0
9	Zakharov	-2.7E+00	1.3E-02	1	-3.8E+00	6.7E-04	1	-1.4E+01	4.0E-14	1	-4.4E+00	1.3E-04	1	nan	nan	0
10	Schwefel 2.22	-4.2E+00	2.4E-04	1	-5.3E+00	1.1E-05	1	-4.8E+01	2.7E-29	1	-1.7E+01	2.7E-16	1	-9.9E+00	9.2E-11	1
11	Schwefel 1.2	-1.4E+00	1.6E-01	0	-1.1E+01	1.2E-11	1	-6.1E+01	3.0E-32	1	-1.1E+01	2.3E-11	1	nan	nan	0
12	Dixon Price	-3.6E+01	1.0E-25	1	-2.7E+02	1.3E-50	1	-8.1E+01	1.1E-35	1	-1.1E+04	7.8E-98	1	-3.0E+04	3.3E-110	1
13	Bohachevsky 1	-2.2E+00	3.3E-02	1	-9.1E+00	5.5E-10	1	-1.3E+01	9.3E-14	1	nan	nan	0	nan	nan	0
14	Booth	5.1E+00	2.1E-05	-1	4.9E+00	3.4E-05	-1	-5.5E+00	7.3E-06	1	5.2E+00	1.6E-05	-1	5.2E+00	1.6E-05	-1
15	Holder Table	-1.8E+01	2.1E-17	1	-1.8E+01	2.1E-17	1	-1.8E+01	2.1E-17	1	-1.8E+01	2.1E-17	1	-1.8E+01	2.1E-17	1
16	Michalewicz 2	-7.1E+02	5.9E-63	1	-6.8E+02	2.0E-62	1	1.4E-59	1.3E-53	-1	-7.1E+02	5.9E-63	1	-7.1E+02	5.9E-63	1
17	Michalewicz 5	-4.5E+01	1.9E-28	1	-4.8E+01	2.9E-29	1	-7.5E+01	9.8E-35	1	-6.9E+01	9.8E-34	1	-7.1E+01	4.4E-34	1
18	Michalewicz 10	-6.7E+01	2.9E-33	1	-6.6E+01	4.2E-33	1	-9.5E+01	9.2E-38	1	-8.0E+01	1.4E-35	1	-7.5E+01	1.0E-34	1
19	Rastrigin	-1.9E+00	7.1E-02	0	-6.9E+00	1.2E-07	1	-4.0E+01	5.0E-27	1	-9.5E+01	9.4E-38	1	-1.0E+01	3.2E-11	1
20	Schaffer 2	4.1E+00	3.1E-04	-1	4.1E+00	3.1E-04	-1	4.1E+00	3.1E-04	-1	4.1E+00	3.1E-04	-1	4.1E+00	3.1E-04	-1
21	Schaffer 4	1.4E+00	1.7E-01	0	1.9E+00	6.7E-02	0	1.8E+00	7.8E-02	0	1.9E+00	6.8E-02	0	1.9E+00	6.6E-02	0
22	Schaffer 6	7.8E+00	1.3E-08	-1	7.8E+00	1.2E-08	-1	-6.4E+01	9.9E-33	1	-6.6E+01	3.1E-33	1	-5.0E+01	8.9E-30	1
23	6HumCame1 Back	-5.9E+04	1.0E-118	1	-6.1E+04	3.5E-119	1	-7.3E+03	2.6E-92	1	-5.9E+04	1.0E-118	1	-5.9E+04	1.0E-118	1
24	Bohachevsky 2	-1.8E+00	7.8E-02	0	-7.8E+00	1.3E-08	1	-1.1E+01	1.5E-11	1	nan	nan	0	nan	nan	0
25	Bohachevsky 3	-1.9E+00	7.0E-02	0	-7.8E+00	1.3E-08	1	-3.8E+00	6.7E-04	1	nan	nan	0	nan	nan	0
26	Shubert	-1.1E+04	1.5E-97	1	-1.1E+04	2.2E-97	1	-1.4E+03	8.3E-72	1	-1.0E+04	2.1E-96	1	-1.1E+04	1.5E-97	1
27	Drop Wave	-inf	0.0E+00	1	-6.6E+07	3.4E-207	1	-3.5E+05	5.8E-141	1	-inf	0.0E+00	1	-inf	0.0E+00	1
28	Rosenbrock	1.7E+03	9.7E-74	-1	-3.3E-02	9.7E-01	0	-5.9E+01	1.0E-31	1	2.2E+01	1.9E-19	-1	8.2E+01	7.4E-36	-1
29	Grienwank	-1.9E+00	7.4E-02	0	-3.8E+00	6.7E-04	1	-1.4E+02	2.2E-42	1	-1.3E+01	6.9E-14	1	nan	nan	0
30	Ackley	-4.9E+00	3.4E-05	1	-1.5E+01	7.1E-15	1	-5.7E+01	2.3E-31	1	-3.2E+01	3.8E-24	1	-7.2E+00	7.0E-08	1

Sign (better) ==> Significance (1 = QMOA; -1 = Alternative Algorithm; 0 = no significant difference), nan = not a number, inf = infinite

TABLE IX. SUMMARY OF SIGNIFICANCE AND RANK

(-1,0,1)	HHO	MS	EHO	A2	A3
QMOA	[5, 8, 17]	[4, 5, 21]	[2, 1, 27]	[3, 6, 21]	[4, 13, 13]

The highest equal performance point, 13 is between QMOA and A3, with QMOA leading with 13 optimal solutions, more than A3's other 4 better performances. Next is HHO with 8 equal points, QMOA with 7 better optimal solutions, and HHO making 5 places. MS and A1 shared very close contest with EHO behind. Table X also is the presentation of the mean runtime measure given below.

TABLE X. RUNTIME TEST RESULTS BASED ON RASTRIGIN

QMOA	HHO	MS	EHO	A2	A3
2.493096	45.6716	756.4435	788.1535	202.0933	217.6285

The performance of QMOA in this section was very high on benchmark complex unconstraint problems compared to contending methods.

H. Results and Statistical Testing with CEC 2017

This section reports QMOA on real-parameter single objective optimization challenging problems featured in Computational Evolution Computation - CEC 2017 with a statistical comparison between QMOA and winners of the competition.

I. Result of QMOA with CEC 2017

The values are the differences between the global optima and the ones obtained with QMOA for 10D, 30D, and 100D during every 51 runs, as shown in the Table XI, and competition is presented in Table XII for 50D

1) 10D, 30D, 50D and 100D Performances

- The uni-modal functions EC1, EC2, and EC3 results were least expected within the number of functional evaluations provided, perhaps due to parameter tuning differences from recommended.

- EC7 – EC10 multimodal functions all attained global optima in all dimensions. QMOA also met 10D and 30D optimal values, with a minor difference for 50D and not too good 100D. EC5 solutions are not good in all dimensions; while EC6 10D was globally optimal, the rest dimensions were not impressive and inadequate for some ranges of solutions.
- Hybrid functions optimization; QMOA yielded optimal global solutions for EC11, EC14 – EC17, and EC19 - EC20 leaving out EC12, EC13, and EC18 with not too good in solutions.
- Besides EC21, EC22, and EC27 of the Composition functions with non-optimal solutions, QMOA achieved optimal global solutions for others, i.e., EC23, EC24, EC25, EC26, EC28, and EC29 in all dimensions, respectively.

J. Time Complexity Analysis

The competition provided appropriate information on the modalities to compute the time complexity [39]. The observation and experimentation shown in Table XII of this work is as follows:

- Evaluate a code consisting of basic arithmetic operation for 1,000,000 iterations and recode the time ( $T_0$ ).
- Evaluate the hybrid function EC18 for 200,000 times the four dimensions (10D, 30D, 50D, and 100D) with record ( $T_1$ ).
- For every dimension, find the meantime of computing  $\bar{T}_2$  the hybrid function EC18 five times run with a termination iteration of 200,000.
- Calculate the time complexity of the algorithm using the relation  $(\bar{T}_2 - T_1) / T_0$ .

Table XIII depicts that the complexity of QMOA is not increasing significantly with the increase in the dimension of the functions.

TABLE XI. STATISTICAL RESULTS FOR CEC 2017 SIMULATIONS D10, D30, AND D100

Tag	Best			Worst			Mean			Median			Standard Deviation		
	10D	30D	100D	10D	30D	100D	10D	30D	100D	10D	30D	100D	10D	30D	100D
EC 1	3.6E+3	6.3E+4	5.8E+5	1.7E+4	1.3E+5	1.7E+6	1.1E+4	9.9E+4	1.1E+6	1.1E+4	1.0E+5	1.1E+6	3.1E+3	1.6E+4	2.7E+5
EC 2	3.6E+3	6.3E+4	5.8E+0	1.7E+4	1.3E+0	1.7E+00	1.1E+4	9.9E+0	1.1E+00	1.1E+4	1.0E+0	1.1E+0	3.1E+3	1.6E+0	2.7E+0
EC 3	1.9E+0	7.3E+4	7.9E+5	1.8E+4	1.9E+5	1.8E+6	9.0E+3	1.3E+5	1.4E+6	8.9E+3	1.3E+5	1.4E+6	3.5E+3	3.1E+4	2.5E+5
EC 4	0.0E+0	0.0E+0	1.9E+3	0.0E+0	0.0E+0	2.5E+3	0.0E+0	0.0E+0	2.3E+3	0.0E+0	0.0E+0	2.3E+3	0.0E+0	0.0E+0	1.4E+2
EC 5	1.5E+7	4.0E+7	1.2E+8	1.5E+7	4.0E+7	1.2E+8	1.5E+7	4.0E+7	1.2E+8	1.5E+7	4.0E+7	1.2E+8	0.0E+0	0.0E+0	0.0E+0
EC 6	0.0E+0	4.4E+2	9.0E+3	0.0E+0	1.8E+3	1.2E+4	0.0E+0	1.3E+3	1.1E+4	0.0E+0	1.3E+3	1.1E+4	0.0E+0	2.9E+2	6.6E+2
EC 7	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 8	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 9	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 10	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0

EC 11	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 12	3.7E+3	4.5E+4	2.4E+5	1.4E+4	7.7E+4	3.2E+5	8.5E+3	6.2E+4	3.0E+5	8.4E+3	6.2E+4	3.0E+5	2.6E+3	6.5E+3	1.5E+4	
EC 13	1.6E+8	1.4E+10	4.9E+10	3.7E+9	3.8E+10	9.4E+10	1.8E+9	2.8E+10	7.3E+10	1.8E+9	2.8E+10	7.4E+10	8.3E+8	5.6E+9	1.1E+10	
EC 14	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 15	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 16	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 17	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 18	3.8E+3	4.6E+4	2.4E+5	1.4E+4	7.7E+4	3.2E+5	8.5E+3	6.2E+4	3.0E+5	8.3E+3	6.2E+4	3.0E+5	2.6E+3	6.5E+3	1.5E+4	
EC 19	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 20	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 21	1.3E+4	1.6E+5	7.9E+5	4.7E+4	2.6E+5	1.1E+6	2.8E+4	2.2E+5	1.0E+6	2.9E+4	2.2E+5	1.0E+6	8.0E+3	2.6E+4	6.8E+4	
EC 22	3.6E+9	2.1E+11	1.8E+12	4.0E+10	6.1E+11	3.3E+12	1.6E+10	4.0E+11	2.7E+12	1.5E+10	3.9E+11	2.7E+12	7.8E+9	9.8E+10	3.3E+11	
EC 23	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 24	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 25	0.0E+0	0.0E+0	5.0E+3	0.0E+0	0.0E+0	6.1E+3	0.0E+0	0.0E+0	5.6E+3	0.0E+0	0.0E+0	5.7E+3	0.0E+0	0.0E+0	2.9E+2	
EC 26	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 27	1.3E+4	1.6E+5	7.9E+5	4.7E+4	2.6E+5	1.1E+6	2.8E+4	2.2E+5	1.0E+6	2.9E+4	2.2E+5	1.0E+6	8.0E+3	2.6E+4	6.8E+4	
EC 28	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
EC 29	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0

TABLE XII. STATISTICAL COMPARISON OF QMOA AND STATE-OF-THE-ART ALGORITHMS FOR CEC 2017, 50D

Tag	JADE	SHADE	UMOEAsII	MVMO	LSHADE-cnEpSin	EBOwithCMAR	QMOA
EC 1	<b>5.2385E-14 (2.5180E-14)</b> +	<b>0.0000E+00 (0.0000E+00)</b> +	<b>0.0000E+00 (0.0000E+00)</b> +	<b>1.3313E-05 (5.6019E-06)</b> +	<b>0.0000E+00 (0.0000E+00)</b> +	<b>0.00+00 (0.00+00)</b> +	2.80E+05 (6.00E+04)
EC 2	1.3112E+13 (8.5354E+13) -	1.0801E+12 (4.3906E+12) -	<b>0.0000E+00 (0.0000E+00)</b> +	1.8060E+17 (1.2778E+18) -	<b>1.5686E+00 (1.9314E+00)</b> +	<b>0.00+00 (0.00+00)</b> +	2.20E+05 (3.60E+04)
EC 3	<b>1.7712E+04 (3.7017E+04)</b> +	<b>0.0000E+00 (0.0000E+00)</b> +	<b>2.1202E-09 (8.8715E-09)</b> +	<b>5.3095E-07 (1.0965E-07)</b> +	<b>0.0000E+00 (0.0000E+00)</b> +	<b>0.00+00 (0.00+00)</b> +	3.40E+05 (6.60E+04)
EC 4	4.9625E+01 (4.7914E+01) -	5.6885E+01 (4.6262E+01) -	6.5462E+01 (5.2164E+01) -	3.5808E+01 (3.6684E+01) -	5.1401E+01 (4.4262E+01) -	4.29E+01 (3.32E+01) -	<b>3.30E+01 (4.80E+01)</b>
EC 5	<b>5.4288E+01 (8.8034E+00)</b> +	<b>3.2859E+01 (5.0387E+00)</b> +	<b>5.0801E+00 (1.6684E+00)</b> +	<b>8.0787E+01 (1.6432E+01)</b> +	<b>2.5166E+01 (6.4447E+00)</b> +	<b>7.58E+00 2.42E+00</b> +	6.60E+07 (0.00E+00)
EC 6	<b>1.4489E-13 (9.1172E-14)</b> +	<b>8.3876E-04 (1.0169E-03)</b> +	<b>1.1951E-06 (1.9013E-06)</b> +	<b>5.4321E-03 (3.3038E-03)</b> +	<b>9.1569E-07 (1.0750E-06)</b> +	<b>8.54E-08 (1.14E-07)</b> +	4.00E+03 (3.80E+02)
EC 7	1.0140E+02 (6.4883E+00) -	8.0964E+01 (3.7800E+00) -	5.6459E+01 (7.1546E-01) -	1.2320E+02 (1.2795E+01) -	7.6639E+01 (6.0618E+00) -	5.79E+01 (1.53E+00) -	<b>0.00E+00 (0.00E+00)</b>
EC 8	5.5234E+01 (7.7643E+00) -	3.2355E+01 (3.8252E+00) -	4.7781E+00 (1.6264E+00) -	7.5910E+01 (1.6122E+01) -	2.6319E+01 (6.5917E+00) -	7.91E+00 (2.47E+00) -	<b>0.00E+00 (0.00E+00)</b>
EC 9	1.1773E+00 (1.3141E+00) -	1.1123E+00 (9.3715E-01) -	1.7555E-03 (1.2536E-02) -	7.3843E+00 (5.7735E+00) -	<b>0.0000E+00 (0.0000E+00)</b> =	<b>0.00+00 (0.00+00)</b> =	<b>0.00E+00 (0.00E+00)</b>
EC 10	3.7500E+03 (2.5448E+02) -	3.3444E+03 (2.9402E+02) -	3.3804E+03 (4.7255E+02) -	3.4971E+03 (4.3138E+02) -	3.2001E+03 (3.3972E+02) -	3.11E+03 (4.01E+02) -	<b>0.00E+00 (0.00E+00)</b>
EC 11	1.3612E+02 (3.3972E+01) -	1.2065E+02 (2.9317E+01) -	4.5701E+01 (9.1852E+00) -	4.7488E+01 (8.7237E+00) -	2.1393E+01 (2.0902E+00) -	2.64E+01 (3.36E+00) -	<b>0.00E+00 (0.00E+00)</b>
EC 12	<b>5.1468E+03 (3.3233E+03)</b> +	<b>5.1362E+03 (2.8785E+03)</b> +	<b>2.1449E+03 (5.3559E+02)</b> +	<b>1.2955E+03 (2.7935E+02)</b> +	<b>1.4753E+03 (3.6472E+02)</b> +	<b>1.94E+03 (5.34E+02)</b> +	1.20E+05 (1.00E+04)

EC 13	<b>3.0338E+02</b> ( <b>2.6999E+02</b> ) +	<b>2.6565E+02</b> ( <b>1.4944E+02</b> ) +	<b>5.1787E+01</b> ( <b>2.1985E+01</b> ) +	<b>4.3776E+01</b> ( <b>1.7622E+01</b> ) +	<b>6.9430E+01</b> ( <b>3.4457E+01</b> ) +	<b>4.14E+01</b> ( <b>2.45E+01</b> ) +	6.90E+10 (1.10E+10)
EC 14	1.0519E+04 (3.1138E+04) -	2.1578E+02 (7.2995E+01) -	2.9299E+01 (2.4831E+00) -	4.8524E+01 (1.2153E+01) -	2.6522E+01 (2.4924E+00) -	3.12E+01 (3.52E+00) -	<b>0.00E+00</b> ( <b>0.00E+00</b> )
EC 15	3.4992E+02 (4.4266E+02) -	3.2262E+02 (1.4201E+02) -	4.1468E+01 (1.0651E+01) -	4.4630E+01 (1.1280E+01) -	2.5596E+01 (4.0567E+00) -	2.94E+01 (5.20E+00) -	<b>0.00E+00</b> ( <b>0.00E+00</b> )
EC 16	8.5696E+02 (1.7532E+02) -	7.3389E+02 (1.8854E+02) -	3.9288E+02 (1.5514E+02) -	8.4082E+02 (1.9349E+02) -	2.7453E+02 (9.9692E+01) -	3.46E+02 (1.46E+02) -	<b>0.00E+00</b> ( <b>0.00E+00</b> )
EC 17	6.0010E+02 (1.2128E+02) -	5.1634E+02 (1.1109E+02) -	3.1356E+02 (1.0636E+02) -	5.1999E+02 (1.3382E+02) -	2.0706E+02 (7.3064E+01) -	2.75E+02 (5.63E+01) -	<b>0.00E+00</b> ( <b>0.00E+00</b> )
EC 18	<b>1.8906E+02</b> ( <b>1.2561E+02</b> ) +	<b>1.8946E+02</b> ( <b>1.0338E+02</b> ) +	<b>3.5997E+01</b> ( <b>8.7118E+00</b> ) +	<b>4.1756E+01</b> ( <b>1.9445E+01</b> ) +	<b>2.4332E+01</b> ( <b>2.1179E+00</b> ) +	<b>3.20E+01</b> ( <b>5.99E+00</b> ) +	1.20E+05 (1.00E+04)
EC 19	3.2429E+02 (1.2561E+03) -	1.5976E+02 (5.6842E+01) -	2.2807E+01 (3.7669E+00) -	1.7338E+01 (5.1321E+00) -	1.7406E+01 (2.4713E+00) -	2.45E+01 (3.94E+00) -	<b>0.00E+00</b> ( <b>0.00E+00</b> )
EC 20	4.3806E+02 (1.3382E+02) -	3.3382E+02 (1.2079E+02) -	2.3041E+02 (1.2312E+02) -	3.2965E+02 (1.4772E+02) -	1.1412E+02 (3.5483E+01) -	1.47E+02 (7.44E+01) -	<b>0.00E+00</b> ( <b>0.00E+00</b> )
EC 21	<b>2.5198E+02</b> ( <b>9.6384E+00</b> ) +	<b>2.3338E+02</b> ( <b>5.1139E+00</b> ) +	<b>2.0681E+02</b> ( <b>2.5498E+00</b> ) +	<b>2.7719E+02</b> ( <b>1.6036E+01</b> ) +	<b>2.2676E+02</b> ( <b>7.0598E+00</b> ) +	<b>2.11E+02</b> ( <b>4.06E+00</b> ) +	4.30E+05 (3.40E+04)
EC 22	<b>3.3364E+03</b> ( <b>1.8053E+03</b> ) +	<b>3.1774E+03</b> ( <b>1.5566E+03</b> ) +	<b>1.7929E+03</b> ( <b>1.9112E+03</b> ) +	<b>3.2653E+03</b> ( <b>1.7185E+03</b> ) +	<b>1.5950E+03</b> ( <b>1.6659E+03</b> ) +	<b>3.65E+02</b> ( <b>9.24E+02</b> ) +	9.20E+11 (1.70E+11)
EC 23	4.7956E+02 (1.1766E+01) -	4.5916E+02 (8.7508E+00) -	4.3459E+02 (5.2143E+00) -	5.0490E+02 (1.5646E+01) -	4.3929E+02 (6.9001E+00) -	4.34E+02 (8.16E+00) -	<b>0.00E+00</b> ( <b>0.00E+00</b> )
EC 24	5.4197E+02 (7.6206E+00) -	5.3106E+02 (7.4577E+00) -	5.0810E+02 (2.6001E+00) -	5.8374E+02 (1.6940E+01) -	5.1282E+02 (5.5948E+00) -	5.06E+02 (3.85E+00) -	<b>0.00E+00</b> ( <b>0.00E+00</b> )
EC 25	5.1923E+02 (3.4820E+01) -	5.0694E+02 (3.6446E+01) -	4.8281E+02 (6.4445E+00) -	5.0912E+02 (3.1226E+01) -	4.8034E+02 (1.0816E+00) -	4.89E+02 (2.47E+01) -	<b>0.00E+00</b> ( <b>0.00E+00</b> )
EC 26	1.6146E+03 (1.2169E+02) -	1.4168E+03 (9.7281E+01) -	5.7211E+02 (4.0709E+02) -	1.9319E+03 (2.8632E+02) -	1.2026E+03 (1.1870E+02) -	7.06E+02 (4.06E+02) -	<b>0.00E+00</b> ( <b>0.00E+00</b> )
F27	<b>5.5080E+02</b> ( <b>2.3427E+01</b> ) +	<b>5.4925E+02</b> ( <b>2.7842E+01</b> ) +	<b>5.3743E+02</b> ( <b>1.7376E+01</b> ) +	<b>5.4355E+02</b> ( <b>1.7557E+01</b> ) +	<b>5.2543E+02</b> ( <b>9.2143E+00</b> ) +	<b>5.22E+02</b> ( <b>7.75E+00</b> ) +	4.30E+05 (3.40E+04)
EC 28	4.9185E+02 (2.0882E+01) -	4.7943E+02 (2.4173E+01) -	4.7289E+02 (2.1643E+01) -	4.6481E+02 (1.5047E+01) -	4.5913E+02 (1.1904E+01) -	4.67E+02 (1.79E+01) -	<b>0.00E+00</b> ( <b>0.00E+00</b> )
EC 29	4.7761E+02 (8.0661E+01) -	4.8716E+02 (1.0502E+02) -	3.6326E+02 (2.0650E+01) -	4.8938E+02 (1.1489E+02) -	3.5289E+02 (9.7796E+00) -	3.47E+02 (1.97E+01) -	<b>0.00E+00</b> ( <b>0.00E+00</b> )
w/t/l	<b>10/0/19</b>	<b>10/0/19</b>	<b>11/0/18</b>	<b>9/0/20</b>	<b>11/1/17</b>	<b>11/1/17</b>	

TABLE XIII. TIME COMPLEXITY ANALYSIS

Time Complexity	T1	T2	TO	(avT2 - T1)/T0
10D	0.164781	0.556989	6.80E-02	5.77E+00
30D	0.1695	0.627578	6.80E-02	6.74E+00
50D	0.18455	0.715318	6.80E-02	7.81E+00
100D	0.24258	0.846715	6.80E-02	8.89E+00

K. Comparison of QMOA with the Winners of CEC2017 (EC1-EC29)

The subsection presents a performance comparison between the 50D problem size for QMOA and other state-of-art high-performing algorithms, especially those that won the CEC2017 competition for real-parameter single objective optimization challenges, as shown in Table XII. The last row represents the values of a Wilcoxon rank-sum test at an alpha value of 0.05. The terms designate the status of the QMOA

against each competing algorithm such that  $w(+.mean..win)/t(=.mean..tie)/l(-.mean..loss)$  For an algorithm making its first entry, the results show very high success [41] shown by Table XII; QMOA had remarkably shown better performance on most of the complex problems considered, as the last row shows. Also, QMOA showed better performances compared with the winners of the competition (competing method, equity, QMOA  $\rightarrow$  w/t/l), e.g., (EBOwithCMAR won 11, equal in 1 and QMOA won 17).

L. Benchmark Design Real Engineering CEC 2020 Single Objective Problems

Eight (8) difficult engineering design-constrained problems that exhibit functional inequality and equality constraints are considered; compared with state-of-the-art algorithms from CEC 2020 real-world optimization issues presented in [42-44]. Among the results presented are the experiments' statistical best, mean, median, worst, and standard deviations. Generally, all models follow a structure as shown in Eq. (15).



$$\min f(x)$$

$$s.t. : g_n(x) \leq 0, n = 1, \dots, m \quad (11)$$

Where  $f$  is the fitness,  $x_s$ ' are the design variables,  $g$  is the constraint with less than equality (often greater than for maximization problems), and  $n$  is the number of constraints. The conversion of the functional constraint from inequality to equality transforms the problem into equation (16).

$$f_p(x) = f(x) + o \sum_{n=1}^m \Phi_n [g_n(x)]^2$$

$$s.t. \quad o > 0 (\text{i.e. penalty factor})$$

$$\Phi_n = \begin{cases} 1 & \text{if } g_n \text{ is violated} \\ 0 & \text{if } g_n \text{ is satisfied} \end{cases} \quad (12)$$

Where  $f_p(x)$  is the penalized objective function. High-performing state-of-the-art algorithms are adopted and compared against the design of certain engineering problems of Fig. 18 (a: Welded beam), (b: Pressure Vessel, and c: Compression Spring). The constraint violations are considered, and the penalty function method is used, which often transforms a constrained problem into an unconstrained continuous counterpart for ease of implementation.

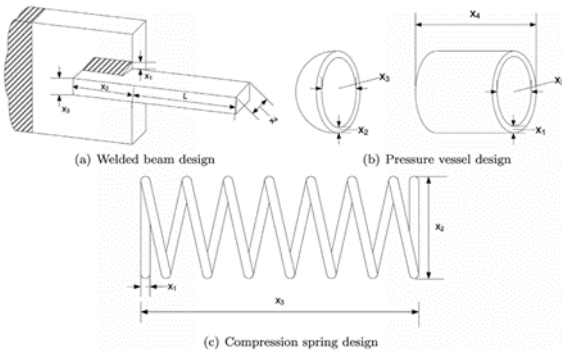


Fig. 18. Engineering design parameter problems.

M. Statistical Comparison of Results for Tension / Compression Spring Design Problem

The design problem in Fig 18 (c) aims at reducing the weight of the tension/compression spring without compromising domain properties like the shear stress, frequency wave, and displacement functionalities [45]. The control variables are wire diameter ( $x_1$ ), mean coil diameter ( $x_2$ ), and the number of coils ( $x_3$ ); the mathematical formulation is detailed in [42]. Upon the experiment, QMOA yielded the most optimal weight compared to the other high-performing algorithms within a minimal number of function evaluations. The result of the compared simulation is shown in Table XIV.

N. Statistical Comparison of the Results for Welded Beam Problem

The welded beam problem Fig. 18 (b) [46] is to minimize the cost of construction. The impacting constraints include shear stress( $\tau$ ); bending stress in the beam ( $\sigma$ ); buckling load

of the bar ( $P_c$ ); end deflection of the beam ( $\delta$ ) and side constraints. The decision variables are (1) the thickness of the weld ( $x_1$ ), the length of the attached part of bar ( $x_2$ ), the height of the bar ( $x_3$ ) and the thickness of the bar ( $x_4$ ). The model formulation is given in [42]. And compared simulated statistical results in Table XV with parametric results in Table XVI.

TABLE XIV. RESULTS FOR THE TENSION / COMPRESSION SPRING DESIGN PROBLEM

Method	Worst	Mean	Best	SD	NFEs
GA1	0.012822	0.012769	0.012704	3.94E-05	900,000
GA2	0.012973	0.012742	0.012681	5.90E-05	80000
CAEP	0.015116	0.013568	0.012721	8.42E-04	50,020
CPSO	0.012924	0.012924	0.012674	5.20E-04	240,000
HPSO	0.012719	0.012707	0.012665	1.58E-05	81,000
NM-PSO	0.012633	0.012631	0.01263	8.47E-07	80,000
G-QPSO	0.017759	0.013524	0.012665	0.001268	2000
QPSO	0.018127	0.013854	0.012669	0.001341	2000
PSO	0.071802	0.019555	0.012857	0.011662	2000
DE	0.01279	0.012703	0.01267	2.7E-05	204,800
DELC	0.012665	0.012665	0.012665	1.3E-07	20,000
DEDS	0.012738	0.012669	0.012665	1.3E-05	24,000
HEAA	0.012665	0.012665	0.012665	1.4E-09	24,000
PSO-DE	0.012665	0.012665	0.012665	1.2E-08	24,950
SC	0.016717	0.012922	0.012669	5.9E-04	25,167
( $\mu + \lambda$ )-ES	NA	0.013165	0.012689	3.9E-04	30,000
ABC	NA	0.012709	0.012665	1.28E-02	30,000
LCA	0.01266667	0.0126654	0.0126652	3.88E-07	15,000
WCA	0.012952	0.012746	0.012665	8.06E-05	11,750
IGMM	0.0135125	0.0128657	0.0126652	2.56E-04	4000
APSO	0.014937	0.013297	0.0127	6.85E-04	120,000
MCEO	0.01350901	0.0127196	0.0126605	3.79E-05	2000
<b>QMOA</b>	<b>0.01160</b>	<b>0.011241</b>	<b>0.011090</b>	<b>0.0002354</b>	<b>2000</b>

TABLE XV. STATISTICAL RESULTS FOR WELDED BEAM PROBLEM

Method	Worst	Mean	Best	SD	NFEs
CAEP	3.179709	1.971809	1.724852	0.443	50,020
CPSO	1.782143	1.748831	1.7314849	0.0129	240,000
HPSO	1.814295	1.74904	1.724852	0.0401	81,000
PSO-DE	1.724852	1.724852	1.724852	6.7E-16	66,600
NM-PSO	1.733393	1.726373	1.72472	0.0035	80,000
SC	6.399678	3.002588	2.385434	0.96	33,095
DE	1.824105	1.768158	1.733461	0.0221	204,800
WCA	1.744697	1.726427	1.724856	0.00429	46,450
LCA	1.7248523	1.7248523	1.7248523	7.11E-15	15,000
IGMM	1.74769	1.732152	1.724855	7.14E-03	8000
APSO	1.993999	1.877851	1.736193	0.076118	50,000
MCEO	1.7248732	1.7248621	1.7248523	1.02E-05	12,500
<b>QMOA</b>	<b>1.6764577</b>	<b>1.622595</b>	<b>1.3534549</b>	<b>0.2390773</b>	<b>60000</b>

TABLE XVI. COMPARISONS OF THE BEST SOLUTIONS FOR WELDED BEAM

DV	CAEP	HGA	NM-PSO	WCA	IGMM	MCEO	QMOA
x1	0.2057	0.2057	0.20583	0.205728	0.205729	0.2057296	<b>0.15203</b>
x2	3.4705	3.4705	3.468338	3.470522	3.470496	3.4704887	<b>8.02626</b>
x3	9.0366	9.0366	9.036624	9.03662	9.036625	9.0366239	<b>6.08668</b>
x4	0.2057	0.2057	0.20573	0.205729	0.205730	0.2057296	<b>0.15775</b>
g1(x)	-769.34	-769.34	-770.3698	-0.034128	-771.187	-771.2021	<b>-0.3760+E4</b>
g2(x)	4.48154	4.48154	-0.053122	-3.49E - 05	-0.05976	-2.88E - 05	<b>-1.83846+E4</b>
g3(x)	-0.2283	-0.2283	-0.228310	-1.19e - 6	-0.228310	-0.228310	<b>-0.00002+E4</b>
g4(x)	0	0	1.00E - 04	-3.43298	-1e - 6	0	<b>0</b>
g5(x)	2.60337	2.603347	-0.031555	-0.080728	-0.0319920	-1.85E - 05	<b>-0.0222+E4</b>
g6(x)	-0.0807	-0.08070	-0.080830	-0.23554	-0.08072	-0.08073	<b>0</b>
g7(x)	-3.4332	-3.43321	-3.43316	-0.013503	-3.4329802	-3.43298	<b>-0.0003+E4</b>
f(x)	1.72457	1.724577	1.724720	1.724856	1.7248552	1.724852	<b>1.622595</b>

In less than 60000 functional evaluations of 10 runs, QMOA yielded a mean cost that is the most optimal for the welded beam in comparison while obeying the constraints.

O. Results for Pressure Vessel Design Problem

The Pressure Vessel Design objective in Fig. 18 (b) [47] is to minimize the cost associated with materials, building, and welding of a cylindrical vessel with capped ends and a hemispherical-shaped head. The impacting variables include the shell thickness x(1), the head thickness x(2), the inner radius x(3), and the length of the cylindrical section excluding the head x(4); the model formulation is given in [42].

The yield of QMOA on the Pressure Vessel design problem produced the best optimal mean value and had a far smaller number of function evaluations of Table XVII.

QMOA met all the inequality constraints; best mean fitness as shown, followed by MCEO, WCA, and NM-PSO, respectively. However, contrary to the large number assigned to the penalty using the other algorithms, QMOA found better results with negligible penalty value for problems of spring and welding beam, and even no penalty was applied to vessel design.

TABLE XVII. THE PRESSURE VESSEL DESIGN PROBLEM RESULTS

Method	Worst	Mean	Best	SD	NFEs
GA1	6308.497	6293.8432	6288.7445	7.4133	900,000
GA2	6469.322	6177.2533	6059.9463	130.9297	80,000
CPSO	6363.8041	6147.1332	6061.0777	86.45	240,000
HPSO	6288.677	6099.9323	6059.7143	86.2	81,000
NM-PSO	5960.0557	5946.7901	5930.3137	9.161	80,000
G-QPSO	7544.4925	6440.3786	6059.7208	448.4711	8000
QPSO	8017.2816	6440.3786	6059.7209	479.2671	8000
PSO	14076.324	8756.6803	6693.7212	1492.567	8000
CDE	6371.0455	6085.2303	6059.734	43.013	204,800
WCA	6590.2129	6198.6172	5885.3327	213.049	27,500
LCA	6090.6114	6070.5884	6059.8553	11.37534	24,000
IGMM	6061.2868	6060.1598	6059.7143	0.5421	8000
APSO	7544.49272	6470.71568	6059.7242	326.9688	200,000
MCEO	6060.3096	6060.0315	6059.7143	1.2532	7500
<b>QMOA</b>	<b>870.8983</b>	<b>848.7333</b>	<b>846.27055</b>	<b>7.7880052</b>	<b>2000</b>

P. Robot Gripper Problem

The complexity involved in manipulating the grippers to minimize the difference between the minimum and maximum forces of the robotic action is ongoing research. Seven design variables, geometric properties, with about seven inequality constraints, are targeted. The Mathematical formulations are found here in [48].

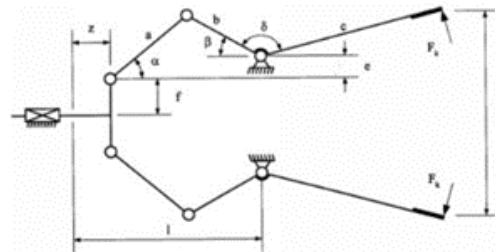


Fig. 19. Schematics of robotics gripper system.

Fig. 19 is schematics of the robotic gripper system, and the experimental result is shown in Table XVIII.

TABLE XVIII. THE STATISTICAL RESULTS OF ROBOTIC GRIPPER OPTIMIZATIONS

Variables	(TLBO)	AOS	QMOA
Best	4.247643634	2.54383686	<b>0.000000000</b>
Mean	4.93770095	2.791745357	<b>0.6661765869</b>
Worst	8.141973	3.14335667	6.4675702717
Std-Dev	0.56	0.226323642	1.8645193071
a	150	149.9973899	27.248574186
b	150	149.880236	150
c	200	200	200
d	0	0	0
e	150	149.9954554	71.18076298
f	100	100.9429469	300
t	2.339539113	2.297394124	2.124666881
g <sub>1</sub> (x)	-28.09283911	-49.99996461	-47.7172
g <sub>2</sub> (x)	-21.90716089	-5.23E-06	-2.2828
g <sub>3</sub> (x)	-33.64959994	-49.99996461	-200.6371
g <sub>4</sub> (x)	-16.35040006	-3.53E-05	150.6371
g <sub>5</sub> (x)	-79999.998	-79737.112	0
g <sub>6</sub> (x)	-9.8E-11	-36.02117726	-4.0000
g <sub>7</sub> (x)	-0.00001	-0.943046876	-0.0200

The experimental result of QMOA on the gripper problem showcases a new optimum as against the optimum global set value [42]; also better than the competing algorithms in comparison [44].

**Q. Rolling Element Bearing**

Five design variables that affect the optimal design of a rolling bearing with the capacity to carry load efficiently amidst nine inequality constraints are considered in the design. The mathematical derivations are provided by [42], while we show the schematics in Fig. 20.

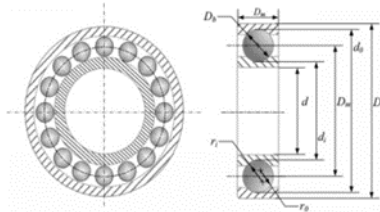


Fig. 20. Schematics of rolling bearing.

The result of the experiment is shown in Table XIX.

TABLE XIX. RESULTS OF ROLLING BEARING AND STATISTIC COMPARISONS WITH QMOA

	(TLBO)	ABC	GWO	ALO	AOS	QMOA
Best	81859.74	85428.24 95	85529.08 30	85546.63 77	83918.492 93	<b>6232.0171 29</b>
Mean	81438.98 7	85121.75 44	83395.08 49	84032.86 36	82175.212 66	<b>9468.3133 94</b>
Worst	80807.85 51	83859.08 51	43543.45 08	73872.81 64	83826.383 37	14246.414 60
Std-Dev	0.66	362.57	8224.5	3121.8	23.38511	2239.4436 27
$D_m$	21.42559	125.6599	125.7090	125.718	125	150
$D_b$	125.7191	21.40862	21.42316	21.42524 2	21.875	10.860427 72
Z	11	11	11	11	10.777009 05	4.510000 00
$f_i$	0.515	0.515	0.515	0.515	0.515	0.5942954 10
$f_o$	0.515	0.515	0.529322	0.515170 18	0.515	0.5802317 89
$K_{Dmi}$	0.424266	0.427166	0.420867	0.454164 6	0.4761106 18	0.4000000 00
$K_{Dma}$	0.633948	0.668849	0.633296	0.646492 4	0.6581426 45	0.6000000 00
E	0.3	0.3	0.300224	0.300001 22	0.3	0.3000000 00
e	0.068858	0.071386	0.02	0.063800 3	0.02	0.0200000 00
Chi	0.799498	0.6	0.619432	0.610759 2	0.6182422 02	0.6000000 00

With an optimum global set at (25287.918415), the experimental result of Table XIX shows that QMOA had set a better global optimum as it also performed better than the competing algorithms [44].

**R. Gas Transmission Compressor Design (GTCD)**

Four variables with one inequality constraint are targeted when designing the gas transmission compressor. The work [42] provides the mathematical formulation while we show the schematics in Fig. 21 and the solutions provided by many optimization state-of-the-art to designs.

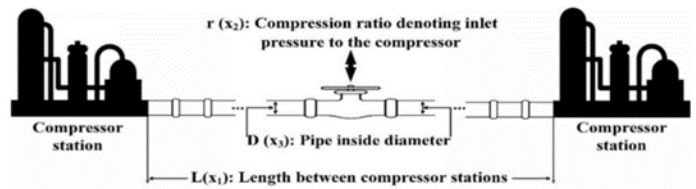


Fig. 21. Schematic of gas transmission compressor system with design variables.

The results of the comparison for the experiment on GTCD are shown in Table XX.

TABLE XX. RESULTS OF OPTIMIZATION OF GTCD AND STATISTICAL OUTCOMES

Algorithms	x1	x2	x3	x4	Optimum cost
CLPSO	45.8830	1.571778	27.18201	1.45592	3.7381430E+06
ABC	50.0000	1.185882	24.89145	0.39507	2.9845610E+06
ACOR	49.6067	1.174456	23.92940	0.37862	2.9671090E+06
ABC	50.0000	1.207839	24.49319	0.45792	2.9755610E+06
KH	35.6206	1.092393	31.99460	1.10937	3.4608480E+06
WOA	49.7095	1.178115	24.72718	0.38796	2.9650350E+06
HHO	49.9844	1.180801	24.20547	0.39429	2.9650910E+06
BOA	20.0000	1.000000	20.00000	0.16475	3.1364520E+06
HGSO	50.0000	1.164785	25.72731	0.35606	2.9689110E+06
LIACOR	50.0000	1.178480	24.58628	0.38882	2.9648960E+06
SMO	50.0000	1.178284	24.59259	0.38835	2.9648954E+06
QMOA	50.0000	1.00000	20.1422	60	<b>9.8081911E+05</b>

The experimental results show QMOA had a set a new global optimum than that set by the competition as the global optimum is (2.9648954173E+06) [42], with the other algorithms as presented in [49].

**S. Himmelblau's Function**

This nonlinear function has been used to test many novel metaheuristic algorithms; it has five main design variables and six inequality constraints to be handled, as shown in Himmelblau [50]. In Table XXI, we show the results of the performances of the metaheuristic algorithms used in comparison.

The experimental result shows that QMOA obtained a better minimum compared to the competing algorithms and set a new global optimum compared to the global presented by [42], which is  $-3.066554E+04$ , with the other algorithms as presented in [49].

**T. Multiple Disk Clutch Brake Design Problem**

The design objective is to minimize the mass of the multiple disk clutch brake, five decision variables with nine nonlinear constraints. The mathematical formulation is given in [42].

TABLE XXI. STATISTICAL AND PERFORMANCE OF ALGORITHMS ON THE HIMMELBLAU COMPLEX PROBLEM

Algorithm	X1	X2	X3	X4	OPTIMAL	
CLPSO	86.3511	34.276	31.279	32.758	-2.99E+04	
ABC	78	33.2729	30.65216	44.30402	36.4902	-3.05E+04
ACOR	78	33	30.04808	44.93806	36.7053	-3.06E+04
ABC	78	33	30.19617	45	36.3524	-3.06E+04
KH	78.99892	33.0057	30.67021	43.63579	35.5313	-3.04E+04
WOA	79.36031	33	30.04906	42.54110	37.2748	-3.05E+04
HHO	78	33	30.00757	44.99297	36.7473	-3.06E+04
BOA	78	33	30.31139	39.59049	31.5837	-3.01E+04
HGSO	78	33	3.109831	4.002299	3.62353	-3.03E+04
LIACOR	78	33	29.99526	45.00000	36.77581	-3.06E+04
SMO	78	33	29.995		36.7758	-3.06E+04
QMOA	79.8729	43.856	27.078	29.1039	29.1039	-3.19E+04

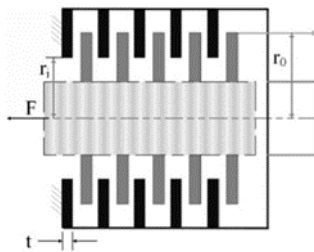


Fig. 22. The schematic geometric representation of the multiple disc clutch design.

Fig. 22 is schematics geometric representation of the clutch disc problem. However, the experimental results are shown in Table XXII.

TABLE XXII. SHOWS THE RESULTS OF THE PERFORMANCE OF METAHEURISTIC METHODS ON THE CLUTCH DESIGN PROBLEM

Algorithms	x1	x2	x3	x4	x5	optimum
CLPSO	75.95932	97.06936	1.01058	909.47864	2.09723	0.280152637613733
ABC	69.99974	90.00000	1.00000	697.47983	2.00000	0.235242474598156
ACOR	70.00000	90.00000	1.00000	718.00397	2.00000	<b>0.235242457900804</b>
ABC	70.00000	90.00000	1.00000	317.17055	2.00000	<b>0.235242457900804</b>
KH	70.00000	90.00000	1.00000	481.07988	2.00000	0.235242458886112
WOA	70.00000	90.00000	1.00000	182.35543	2.00000	0.235242457901052
HHO	70.00000	90.00000	1.00000	304.20738	2.00000	<b>0.235242457900804</b>
BOA	67.72699	90.00000	1.00000	673.06921	2.00000	0.248171278270212
HGSO	69.99945	90.00000	1.00000	8.73600	2.00000	0.235248138956563
LIACOR	70.00000	90.00000	1.00000	169.99845	2.00000	<b>0.235242457900804</b>
SMO	70.000	90.000	1.000	999.99	2.000	<b>0.2352424579</b>
QMOA	80.000	90.005	1.000	1000.0	2.000	<b>0.1250434142</b>

QMOA performed better than ABC, which was reported as best at the time of competition report, and others in this design problem and further set a much better global optimum than benchmarked in [42]; 0.23524245790; with the other algorithms as presented in [49].

#### IV. CONCLUSION

In this work, QMOA, a new nature-inspired population-based metaphor, was proposed and used in experiments and engineering designs with very great performance. The idea stemmed from the informal learning pattern and discipleship, which is ingrained in the socio-cultural behavior of the indigenous peoples - the Ndigbo of a West African tribe is presented. The learners cope through practice and observation. The experiment conducted considered 30 benchmark unconstrained problems, 29 CEC 2017 (50D) real-parameter single objective constraint optimization, and about 8 engineering design constrained problems from CEC 2020; the results showed that QMOA had balanced exploitation and exploratory capacities with very good convergence time too. Comparison to the performance of other well-established state-of-the-art algorithms depicts the exceptional performance of the automata. The significant test also confirms the relative efficiency of QMOA with t-values and p-values presented in Table VIII and summarized in Table IX. The convergence time test using the Rastrigin function also shows QMOA had better speed than the contender in Table X. The competing algorithms were the most award winners in past competitions from 2017 till date. In all complex engineering problems presented, QMOA had performed remarkably well and had, in some cases, set new minimum attainable best solutions; Of interest are the new values better than the set global optimums in some functions and engineering designs (Clutch Disc, Himmelblau, GTCD, Rolling Bearing, Robotic Gripper).

The future direction is to further validate with the most recent CECs and design optimization problems in other fields. Meanwhile, QMOA shows merit to be considered in the current state-of-the-art.

#### ACKNOWLEDGMENT

Akowuah E. K.: Lead Supervision, Validation, Revising the manuscript, providing suggestions, and Feedback on the methodology. Kponyo J. J.: Supervision, Revising the manuscript, Providing suggestions, and Feedback on the methodology. Boateng K. O.: Supervision, Providing Feedback on the approach.

#### REFERENCES

- [1] B. A. Bhuiyan, "An overview of game theory and some applications," *Philosophy and Progress*, vol. 59, pp. 111-128, 2018.
- [2] A. Vasuki, *Nature-Inspired Optimization Algorithms*: CRC Press, 2020.
- [3] K. Sorensen, M. Sevaux, and F. Glover, "A history of metaheuristics," arXiv preprint arXiv:1704.00853, 2017.
- [4] A. P. Engelbrecht, *Computational intelligence: an introduction*: John Wiley & Sons, 2007.
- [5] D. Karaboga and B. Akay, "A comparative study of artificial bee colony algorithm," *Applied mathematics and computation*, vol. 214, pp. 108-132, 2009.

- [6] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE transactions on evolutionary computation*, vol. 1, pp. 67-82, 1997.
- [7] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*: MIT press, 1992.
- [8] R. Eberhart and J. Kennedy, "Particle swarm optimization," in *Proceedings of the IEEE international conference on neural networks*, 1995, pp. 1942-1948.
- [9] S. J. Ceci, *On intelligence*: Harvard University Press, 1996.
- [10] D. Teodorović, "Bee colony optimization (BCO)," in *Innovations in swarm intelligence*, ed: Springer, 2009, pp. 39-60.
- [11] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE computational intelligence magazine*, vol. 1, pp. 28-39, 2006.
- [12] H. Missbauer and R. Uzsoy, "Optimization models of production planning problems," in *Planning production and inventories in the extended enterprise*, ed: Springer, 2011, pp. 437-507.
- [13] F. Glover, "Tabu search—part II," *ORSA Journal on computing*, vol. 2, pp. 4-32, 1990.
- [14] H. Eskandar, A. Sadollah, A. Bahreininejad, and M. Hamdi, "Water cycle algorithm—A novel metaheuristic optimization method for solving constrained engineering optimization problems," *Computers & Structures*, vol. 110, pp. 151-166, 2012.
- [15] A. Kaveh and T. Bakhshpoori, "A new metaheuristic for continuous structural optimization: water evaporation optimization," *Structural and Multidisciplinary Optimization*, vol. 54, pp. 23-43, 2016.
- [16] M. Azizi, "Atomic orbital search: A novel metaheuristic algorithm," *Applied Mathematical Modelling*, vol. 93, pp. 657-683, 2021.
- [17] M. A. Shaheen, H. M. Hasanien, and A. Alkuhayli, "A novel hybrid GWO-PSO optimization technique for optimal reactive power dispatch problem solution," *Ain Shams Engineering Journal*, vol. 12, pp. 621-630, 2021.
- [18] H. AlSattar, A. Zaidan, B. Zaidan, M. Abu Bakar, R. Mohammed, O. Albahri, et al., "MOGSABAT: a metaheuristic hybrid algorithm for solving multi-objective optimisation problems," *Neural Computing and Applications*, vol. 32, pp. 3101-3115, 2020.
- [19] X. S. Yang and A. H. Gandomi, "Bat algorithm: a novel approach for global engineering optimization," *Engineering computations*, 2012.
- [20] A. Gogna and A. Tayal, "Metaheuristics: review and application," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 25, pp. 503-526, 2013.
- [21] W. Wong and C. I. Ming, "A review on metaheuristic algorithms: recent trends, benchmarking and applications," in *2019 7th International Conference on Smart Computing & Communications (ICSCC)*, 2019, pp. 1-5.
- [22] O. Adeola, "The Igbo Business Practice: Towards a Model for Africa Conclusion and Recommendations," in *Indigenous African Enterprise*, ed: Emerald Publishing Limited, 2020.
- [23] O. A. Olutayo, "The Igbo entrepreneur in the political economy of Nigeria," *African Study Monographs*, vol. 20, pp. 147-174, 1999.
- [24] T. Ubesie, "Odinala Ndi Igbo," *Ibadan: Oxford University*, vol. 999, 1978.
- [25] U. U. David, "Chinua Achebe and Flora NWAPA at the Biafran Literary War Front."
- [26] I. O. Iwara, K. E. Amaechi, and V. Netshandama, "The Igba-boi apprenticeship approach: Arsenal behind growing success of Igbo entrepreneurs in Nigeria," *African Journal of Peace and Conflict Studies*, pp. 227-250, 2019.
- [27] C. C. Kanu, "The context of Igwebuike: What entrepreneurship development systems in Africa can learn from the Igbo apprenticeship system," *AMAMIHE Journal of Applied Philosophy*, vol. 18, 2020.
- [28] B. Bhushan, *Principles and applications of tribology*: John Wiley & Sons, 1999.
- [29] J. A. LePine and L. Van Dyne, "Voice and cooperative behavior as contrasting forms of contextual performance: evidence of differential relationships with big five personality characteristics and cognitive ability," *Journal of applied psychology*, vol. 86, p. 326, 2001.
- [30] C. Omeire, E. Omeire, P. Nwaoma, A. Otunko, and P. Onoh, "THE BIAFRA QUESTION: ASocio-CULTURAL EXAMINATION OF THE IGBO NATION OF SOUTH EASTERN NIGERIA," *International Journal of Social Sciences, Humanities and Education*, vol. 1, pp. 1-9, 2017.
- [31] J. K. Osiri, "Igbo management philosophy: A key for success in Africa," *Journal of Management History*, 2020.
- [32] A. J. Nebro, J. J. Durillo, J. Garcia-Nieto, C. C. Coello, F. Luna, and E. Alba, "SMP SO: A new PSO-based metaheuristic for multi-objective optimization," in *2009 IEEE Symposium on computational intelligence in multi-criteria decision-making (MCDM)*, 2009, pp. 66-73.
- [33] H. Meunier, E.-G. Talbi, and P. Reininger, "A multiobjective genetic algorithm for radio network optimization," in *Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No. 00TH8512)*, 2000, pp. 317-324.
- [34] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," *Future generation computer systems*, vol. 97, pp. 849-872, 2019.
- [35] G.-G. Wang, "Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems," *Memetic Computing*, vol. 10, pp. 151-164, 2018.
- [36] G.-G. Wang, S. Deb, and L. d. S. Coelho, "Elephant herding optimization," in *2015 3rd international symposium on computational and business intelligence (ISCBI)*, 2015, pp. 1-5.
- [37] A. W. Mohamed, A. A. Hadi, A. M. Fattouh, and K. M. Jambi, "LSHADE with semi-parameter adaptation hybrid with CMA-ES for solving CEC 2017 benchmark problems," in *2017 IEEE Congress on evolutionary computation (CEC)*, 2017, pp. 145-152.
- [38] A. Kumar, R. K. Misra, and D. Singh, "Improving the local search capability of effective butterfly optimizer using covariance matrix adapted retreat phase," in *2017 IEEE congress on evolutionary computation (CEC)*, 2017, pp. 1835-1842.
- [39] G. Wu, R. Mallipeddi, and P. N. Suganthan, "Problem definitions and evaluation criteria for the CEC 2017 competition on constrained real-parameter optimization," *National University of Defense Technology, Changsha, Hunan, PR China and Kyungpook National University, Daegu, South Korea and Nanyang Technological University, Singapore*, Technical Report, 2017.
- [40] A. Ghosh, S. Das, and A. K. Das, "A simple two-phase differential evolution for improved global numerical optimization," *Soft Computing*, vol. 24, pp. 6151-6167, 2020.
- [41] R. Salgotra, U. Singh, S. Saha, and A. H. Gandomi, "Improving cuckoo search: incorporating changes for CEC 2017 and CEC 2020 benchmark problems," in *2020 IEEE Congress on Evolutionary Computation (CEC)*, 2020, pp. 1-7.
- [42] A. Kumar, G. Wu, M. Z. Ali, R. Mallipeddi, P. N. Suganthan, and S. Das, "A test-suite of non-convex constrained optimization problems from the real-world and some baseline results," *Swarm and Evolutionary Computation*, vol. 56, p. 100693, 2020.
- [43] F. MiarNaeimi, G. Azizyan, and M. Rashki, "Multi-level cross entropy optimizer (MCEO): an evolutionary optimization algorithm for engineering problems," *Engineering with Computers*, vol. 34, pp. 719-739, 2018.
- [44] M. Azizi, S. Talatahari, and A. Giaralis, "Optimization of engineering design problems using atomic orbital search algorithm," *IEEE Access*, vol. 9, pp. 102497-102519, 2021.
- [45] A. D. Belegundu and J. S. Arora, "A study of mathematical programming methods for structural optimization. Part I: Theory," *International Journal for Numerical Methods in Engineering*, vol. 21, pp. 1583-1599, 1985.
- [46] C. A. C. Coello, "Use of a self-adaptive penalty approach for engineering optimization problems," *Computers in Industry*, vol. 41, pp. 113-127, 2000.
- [47] A. H. Gandomi, X.-S. Yang, and A. H. Alavi, "Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems," *Engineering with computers*, vol. 29, pp. 17-35, 2013.
- [48] R. V. Rao and G. Waghmare, "Design optimization of robot grippers using teaching-learning-based optimization algorithm," *Advanced Robotics*, vol. 29, pp. 431-447, 2015.

- [49] H. Zamani, M. H. Nadimi-Shahraki, and A. H. Gandomi, "Starling murmuration optimizer: A novel bio-inspired algorithm for global and engineering optimization," *Computer Methods in Applied Mechanics and Engineering*, vol. 392, p. 114616, 2022.
- [50] D. M. Himmelblau, *Applied nonlinear programming*: McGraw-Hill, 2018.

# Modeling of Whale Optimization with Deep Learning based Brain Disorder Detection and Classification

Uvaneshwari M<sup>1</sup>, Dr. M. Baskar<sup>2\*</sup>

Research Scholar-Department of Computer Science and Engineering-School of Computing-College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamilnadu, 603 203, India<sup>1</sup>  
Associate Professor-Department of Computing Technologies-School of Computing-College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamilnadu, 603 203, India<sup>2</sup>

**Abstract**—Brain disorders are a significant source of economic strain and unfathomable suffering in modern society. Imaging techniques help diagnose, monitor and treat mental health, neurological, and developmental disorders. To aid in the Computer-Aided Diagnosis of brain diseases, deep learning (DL) was used for the analysis of neuroimages from modalities including Positron Emission Tomography (PET), Structural Magnetic Resonance Imaging (SMRI), and functional MRI. In this study, a Whale Optimization Algorithm is used with Deep Learning to analyse MRI scans for signs of neurological disease. WOADL-BDDC may detect and label abnormalities in the brain based on an MRI scan. It uses a two-step pre-processing procedure, first using guided filtering to get rid of background noise and then using U-Net segmentation to get rid of the top of the head. QuickNAT, along with RMSProp, is used to segment the brain. When analysing data, WOADL-BDDC uses radionics to collect information from every layer. When used in a convolutional recurrent neural network model, the Whale Optimization Algorithm may accurately categorize mental illness. WOADL-BDDC is put through its paces using ADNI 3D. Compared to state-of-the-art classification results from Vgg16, Graph CNN, Modified ResNet18, Non-linear SVM, ResNet50-SVM, ResNet50-RF, the suggested technique achieved the greatest accuracy. It was demonstrated that the suggested model is superior to other models for classification from MRI images. In simulations, the proposed approach is shown to be effective in optimizing hyperparameters with an accuracy of 94.38 % on TR set and 94.87% on TS set, Precision of 96.43% on TR set and 97.62% on TS set, and an F1-Score of 89.35 % and 92.10% on TR and TS set, respectively.

**Keywords**—Brain disorder detection; magnetic resonance imaging; deep learning; convolutional recurrent neural network; whale optimization algorithm

## I. INTRODUCTION

The brain is the management center of the central nervous system and is liable for implementing activities all over the human body [1, 2]. As in other medical sectors, there exist innovative neuroimaging technologies comprising positron emission tomography (PET), and MRI was introduced and utilized for identifying Alzheimer's Disease (AD) based molecular and structural biomarkers [3,17]. Rapid progression in neuroimaging techniques makes them challenging to incorporate largescale, higher-dimension multimodal neuroimaging datasets [4]. Thus, computer-aided machine learning (ML) algorithms for integrative analysis have gained considerable interest. To apply this ML approach, Pre-

Processing steps or appropriate architectural design must be defined in advance [5]. Generally, a classification study using the ML method needs four stages: feature selection, feature extraction, feature-based classification algorithm selection, and dimensionality reduction [6]. This process requires multiple optimization steps and professional knowledge that might take considerable time [7]. To conquer this problem, deep learning (DL), an emerging field of ML study that uses raw neuroimaging information to generate features through "on-the-fly" learning, is received significant interest in the Area of large-scale, higher-dimension medical imaging analysis. DL approaches, like convolution neural networks (CNN), have proved to outperform the present ML method [8, 9]. Compared with conventional ML techniques, the DL approach automatically discovers the proper representation without specialized knowledge and permits the non-expert to utilize DL approaches efficiently. Thus, the DL method has rapidly become a method of choice for medical image analysis [10].

This study progresses a Whale Optimization Algorithm with Deep Learning based Brain Disorder Detection and Classification (WOADL-BDDC) on MRI images. The WOADL-BDDC technique initially performs pre-processing in 2 stages: guided filtering (GF) based noise elimination and U-Net segmentation-based skull stripping. Besides, the QuickNAT with RMSProp optimizer is applied to segment the brain parts. For feature extraction, the WOADL-BDDC technique performs a radionics approach that separates features of all aspects. Finally, the WOA with convolutional recurrent neural network (CRNN) model is applied to classify brain disorders. The experimental evaluation of the WOADL-BDDC technique is tested on the ADNI 3D dataset.

The wide range in tumour size, shape, and location poses a significant problem for diagnosing brain tumours. The scientific community can benefit from this review since it provides a thorough literature on brain tumour detection using magnetic resonance imaging. The healthcare sector is unlike any other. Regardless of financial constraints, this is a top priority area that attracts customers expecting only the best service. Given deep learning's track record of success in the real world, it should be no surprise that this promising technique is already playing a crucial role in medical imaging, yielding promising results with high accuracy. The brain is the master organ, regulating the functions of every other body component.

\*Corresponding Author.

Due to the intricacy of size and position variations, automated brain tumour recognition in MRI is challenging. The pictures produced from MRI scans of the brain are processed using proposed statistical, morphological, and thresholding methods to detect tumours. Based on our findings, we can confidently state that conventional algorithms perform admirably for both the number of initial clusters and the locations of their centers. Pixel classification becomes difficult if these groups shift depending on the parameters you use to construct them. As it is, the cluster centroid value is chosen at random in the most common fuzzy cluster mean procedure. Because of this, it will take longer to get the intended result. Radiologists have to spend a lot of time manually segmenting and evaluating MRI brain pictures. The segmentation is performed using machine learning approaches that are less accurate and have a slower calculation speed. Tumor categorization and identification using neural network methods have been attempted extensively but with mixed results. Detection accuracy relies on the segmentation and detection algorithms employed. Thus far, current systems need to provide better picture quality and precision.

A convolutional recurrent neural network (CRNN) will be employed to categorize the tumors in the picture. Increased precision and decreased repetition are the only two benefits of this method.

## II. LITERATURE REVIEW

The authors [11] modeled a deep ensemble learning structure to harness DL techniques to tap the 'wisdom of experts' and compile multisource data. At the voting layer, two sparse AEs were instructed for feature learning to minimize the correlation of features and expand the base classifier eventually. At the stacking layer, a non-linear feature-weighted technique related to a DBN was devised to rank the base classifiers, which disrupt the conditional independence. In [12], the authors intend to formulate an innovative DL technique for detecting or productively predicting AD. The authors devised densely connected CNNs with a connection-wise attention system for learning the multilevel attributes of brain MRI for AD categorization. The authors employed the densely associated NN for deriving multiscale features in pre-processing images. A connection-wise attention system has been implemented for integrating associations among attributes from distinct layers. In [13], the authors employ a DL-related Granger causality estimation of brain connectivity structure. It uses the robustness of LSTM in changing time series processing. The authors employ MRI to examine the cerebral cortex property and exploit rs-fMRI to analyze the function network's graph metric. In [14], the authors establish that it is settled well with the multi-task learning method. In [15] devised a DL technique for every level of feature extracting process and fuzzy hyper-plane related least square twin SVM (FLS-TWSVM) for classifying the derived features for initial analysis of AD through removing sagittal plane slices from 3-D MRI imageries. In [16], an Alzheimer's classification and detection method was provided. The BoVW method was employed to enhance the efficacy of texture-related features, like the histogram of gradient, GLCM, and scale-invariant feature transform, LBP.

Brain tumours are notoriously difficult to diagnose because of the wide range of tumour sizes, forms, and shapes that must be considered during the process [25]. Surgical, radiological, and chemotherapeutic options are all on the table when a brain tumour diagnosis has been made, and prognoses can be formulated based on the tumor's nature and location [26]. Therefore, early diagnosis and identification of brain tumours aid in treatment planning and patient monitoring. It's a vital factor in better treatment and increased survival rates. To learn more about tumours, many medical imaging and diagnostic methods are performed. It is possible to distinguish between usually and abnormally expanding brain cells with the aid of imaging techniques like magnetic resonance imaging (MRI) and computed tomography (CT)[27]. Computed Tomography (CT) scans are utilized for diagnosis through X-ray and computer to produce pictures of the patient's brain in axial slices [28].

According to the systematic literature evaluation, there is space for improvement in brain tumour detection. Because brain tumours can be any size or shape, current segmentation methods need further work before they can be used for tumours. The importance of enhancement and segmentation in tumour identification stems from their role in addressing the shortcomings of current approaches.

The significance of medical data offered along with imaging data can be emphasized by incorporating clinical features, including texture-related features, for generating hybrid feature vectors. Though several computer-aided diagnosis models are existed in the literature to perform the brain tumor classification processes, it is required to improve the classifier results. Due to the repeated extension of the model, the number of parameters of DL models also rises rapidly, which leads to model overfitting. So, the WOA algorithm is used to select CRNN parameters. Personalized image collection based on estimates has been proven to be capable of seeing channels during a CT examination, thanks to the ongoing development of AI. Significant learning, in particular, has demonstrated encouraging outcomes in automated planning. It's so good at processing visual information that we can now employ neural connections to construct meaningful images rather than merely sort them. In particular, CRNN Association is well suited for evaluating images from channels like CT and X-ray. CRNNs are mainly designed to handle photographs better and finish picture groups. In this way, CRNNs can approach radiologists' accuracy when recognizing crucial elements in CT channels or other signature images.

## III. THE PROPOSED WOADL-BDDC MODEL

In this study, we have established a new WOADL-BDDC technique for recognizing brain disorders on MRI images. The WOADL-BDDC approach encompasses GF-based noise removal, U-Net-based skull stripping, QuickNAT with RMSProp-based brain segmentation, radionics feature extraction, CRNN classification, and WOA hyperparameter tuning. Fig. 1 represents the block diagram of the WOADL-BDDC system.



### A. Image Pre-Processing

A GF is an edge-preserving smoothing light filter. It applies edge-preserving smoothing on the image, utilizing the content of the secondary image, named a guidance image, to impact the filtering. The guidance image could be a completely different image or different versions of an image. The bilateral filter might filter out texture or noise when maintaining sharp edges. A Gaussian operator can compute the convolution operator, and a proposal of Gaussian smoothening can be made by convolution. The Gaussian operator can be defined using the following equations:

$$G_{1D}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x^2}{2\sigma^2}\right)} \quad (1)$$

The optimum smoothening filter for the image can be localized in the frequency and spatial domains, where the ambiguity relativity can be satisfied as follows [19]:

$$\Delta x \Delta \omega \geq \frac{1}{2}. \quad (2)$$

The Gaussian operator in 2D can be represented by:

$$G_{2D}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \quad (3)$$

Where  $\sigma$  (sigma) signifies the standard deviation (SD) of Gaussian and  $(x, y)$  represents the Cartesian coordinates of the image. In medical applications, it serves as an initial phase since it identifies path abnormality in the speed, brain, and accuracy of diagnoses. It isolates non-brain tissues such as the scalp and skull from the brain image. Skull stripping removes unwanted components and non-brain anatomy from scanned images. There exist five convolution blocks in the downsampling route. Currently, there are overall of 1024 feature maps. Excepting the last block, max pooling with stride 22 can be performed after the completion of every block for downsampling. The feature map is reduced in size from 240x240 to 1515. Each up-sampling block initiates by a deconvolution layer of stride 22 and filters size 33. Consequently, feature map is becoming increasingly popular. The two convolution layers in the up-sampling block lower the quantity of deconvolution feature map and feature map in the encoding route.

### B. Brain Segmentation

In this work, the QuickNAT carries out the segmentation of brain regions with RMSProp optimizer. QuickNAT is the first method that uses an unlabelled neuroimaging dataset with auxiliary labels for training and a higher testing speed of the FC network for brain segmentation [18]. QuickNAT comprises sagittal, coronal, and axial views monitored by a view aggregation stage to deduce the last segment. Every FCNN is a similar structure and is stimulated by the conventional encoding/decoding-based UNet structure with skip connection improved with an un-pooling layer. Also, the study presents dense links in all the encoding/decoding blocks to promote feature re-usability and to help gradient flow, that is, the small number of training datasets. A multi-class Dice loss and weighted logistic loss can enhance the network.

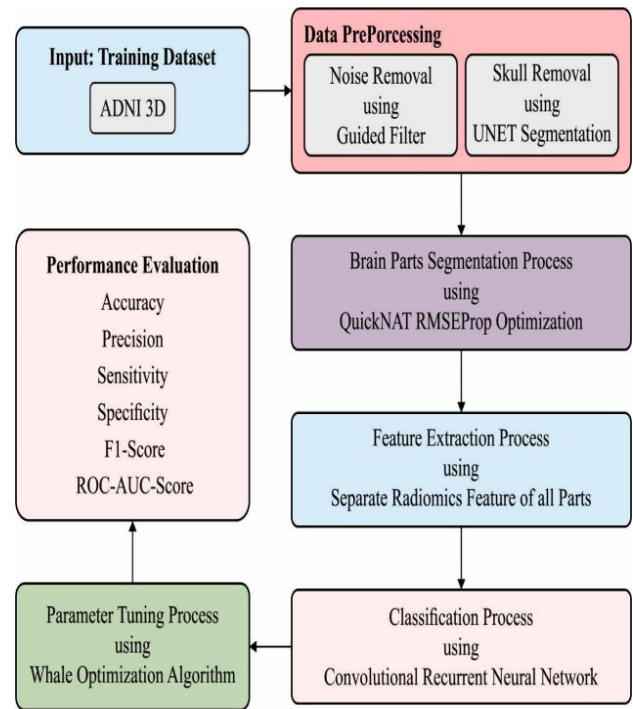


Fig. 1. Block diagram of WOADL-BDDC system.

In contrast, it compensates for higher-class imbalanced datasets and encourages the accurate assessment of anatomical boundaries using a proper weighting system. The two primary techniques of QuickNAT are the training method with auxiliary labels and the FCNN framework. The trained model and code are presented as extensions of MatConvNet for off-the-shelf usage.

### C. Radiomics Feature Extraction

For the process of extracting features, the presented algorithm will use radiomics features. The abovementioned features are derived by in-house software, utilizing PyRadiomics and python's skit-learn package [19]. Texture, intensity, and shape are the three kinds of features that are calculated. Twenty-six shape-based and intensity-based features can be extracted for all extraction settings (width, filter, bin, and contour). With sixteen distinct stages, which can be permutations of 2 contours (HGG and Gold), bandwidth (2, 4, 8, 16), and two filters (LoG and original) and the entire count of derived radiomics imaging features was 1450.

### D. CRNN-Based Brain Disorder Classification

The CRNN model is applied here to detect and classify brain disorders. The study presents the CRNN utilized for organizing MRI [20]. Initially, CNN is used for learning higher-level feature representation. Next, the CNN-learned feature was fed into the Bi-GRU layer for learning the temporal correlation data. At last, this feature was provided in the FC layer by Softmax function for the outcome of the likelihood distribution of distinct classes. In the presented method, the convolution RNN has encompassed two BiGRU layers ( $l_9 - l_{10}$ ) and eight convolution layers ( $l_1 - l_8$ ). The network parameter and architecture are given below:

$l_1 - l_2$ : The first two stacked convolution layer uses 32 filters with the stride of (1,1) and receptive field of (3,5). Then, max-pooling with the stride of (4,3) decreases the size of feature maps.

$l_3 - l_4$ : The subsequent two convolution layers use 64 filters with a stride of (1,1) and a receptive field of (3,1) and are utilized for learning local patterns alongside the frequency dimension. It can be max-pooling with the stride of (4,1).

$l_5 - l_6$ : The succeeding pairs of convolution layer applied 128 filters with a stride of (1,1) and receptive field of (1,5) and are utilized for learning local patterns along the time dimension. It can subsequently be the max-pooling with the stride of (1,3).

$l_7 - l_8$ : The successive two convolution layer uses 256 filters with a stride of (1,1) and receptive field of (3,3) for learning joint time–frequency features. It can subsequently be the max-pooling of (2,2) stride.

$l_9 - l_{10}$ : 2BiGRU layers with 256 cells were utilized for temporal summarization, and the activation function was used. The dropout with the possibility of 0:5 is applied to every BiGRU layer to prevent over-fitting. Fig. 2 illustrates the framework of CRNN.

#### E. Hyperparameter Tuning

For adjusting the hyperparameters related to the CRNN model, the WOA is utilized in this work. WOA is a population-based heuristic approach that models the hunting strategy of humpback whales [21]. They construct unique maneuvers for paralyzing the prey by swimming nearby them in a spiral or circle-shaped path. This technique is named the bubble-net attack. The hunting behavior can be specified into two stages. The initial stage is the exploration phase, where the whales randomly seek prey. The next stage is the exploitation stage, whereby the whales implement bubble-net hunting. The following equation is utilized for mathematically modeling these behaviors as follows:

$$\vec{D} = |\vec{C} \cdot \vec{X}(t) - \vec{X}(t)| \quad (4)$$

$$\vec{X}(t+1) = \vec{X}(t) - \vec{A} \cdot D \quad (5)$$

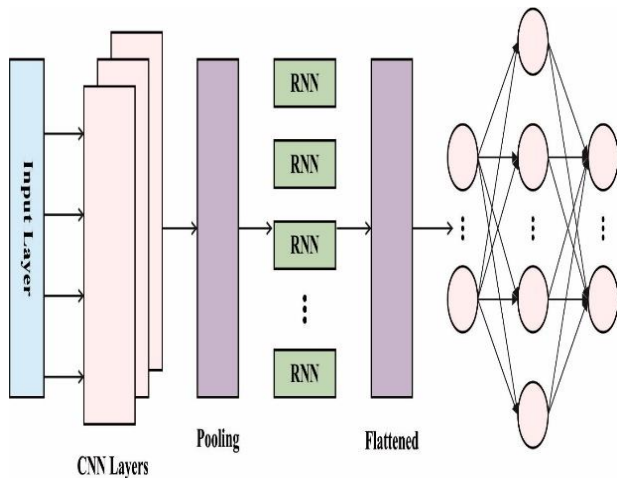


Fig. 2. Framework of CRNN.

Whereas  $\vec{X}^*$  represents the best position of the attained solution,  $t$  characterizes the count of iterations, and the constant variables  $\vec{A}$  and  $\vec{C}$  are characterized by:

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad (6)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (7)$$

Now,  $\vec{r}$  denotes an arbitrary vector amongst [0, 1], and variable  $\vec{a}$  is an arbitrary vector utilized for controlling the convergence procedure. It can be linearly decreased from two to zero through iteration:

$$\vec{a} = 2 - t \frac{2}{T} \quad (8)$$

In Eq. (5),  $t$  indicates the present iteration, and  $T$  shows the maximal iteration count of the vector  $\vec{A}$  utilized for displaying the transition method amongst the exploratory and exploitative actions.

In WOA, the exploitation stage could be one of 2 models, shrinking surroundings or spiral-shaped methods. During the shrinking surrounding model, it is accomplished by reducing the value. The spiral path model upgrades the distance betwixt the most efficient search agents attained up to this point ( $\vec{X}^*$ ) and searching agents ( $X$ ), then boosts a novel location for the searching agent based on the subsequent as:

$$X(t+1) = D \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \quad (9)$$

The probabilities value of 0.5 is set as follows:

$$\vec{X}(t+1) = \begin{cases} D' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t), & p \geq 0.5 \\ \vec{X}(t) - \vec{A} \cdot D, & p < 0.5 \end{cases} \quad (10)$$

Whereas  $p$  denotes an arbitrary probability value. The pseudocode of the proposed model is shown in Algorithm 1.

Algorithm 1: Overall process of WOADL-BDDC technique
<b>Input:</b> ADNI 3D Dataset
<b>Output:</b> Classified images
<b>Start</b>
For every image
do
<b>Image_Pre-processing</b> (GF_noise, U_Net skull stripping)
<b>Segment_image()</b>
QuickNAT mode
Hyperparameter adjustment using RMSProp Optimizer
<b>End</b>
<b>Feature_extraction(radiomics)</b>
Determine texture, shape, intensity
<b>Classify()</b>
Execute CRNN
Hyperparameter Tuning using WOA
Execute Exploration stage
Execute Exploitation stage
<b>End</b>
<b>End</b>

IV. RESULTS AND DISCUSSION

The presented WOADL-BDDC system is simulated using a Python tool. The results are examined under the ADNI 3D dataset [22].

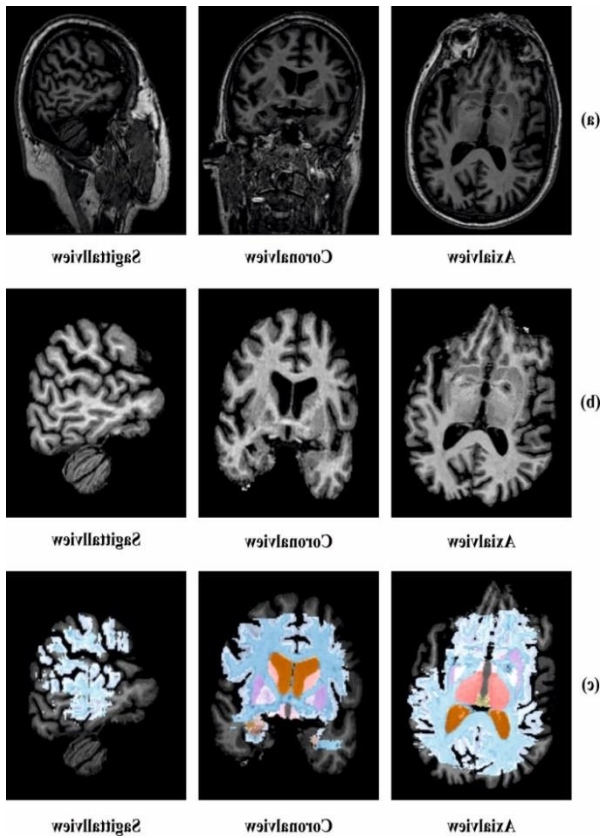


Fig. 3. (a) Original, (b) Pre-processed, (c) Segmented.

Fig. 3 visualization of the sample outcomes offered by the WOADL-BDDC system. Fig. 3(a) depicts the original brain MRI and the pre-processed image is provided in Fig. 3(b). Besides, the segmented images by the WOADL-BDDC system are shown in Fig. 3(c).

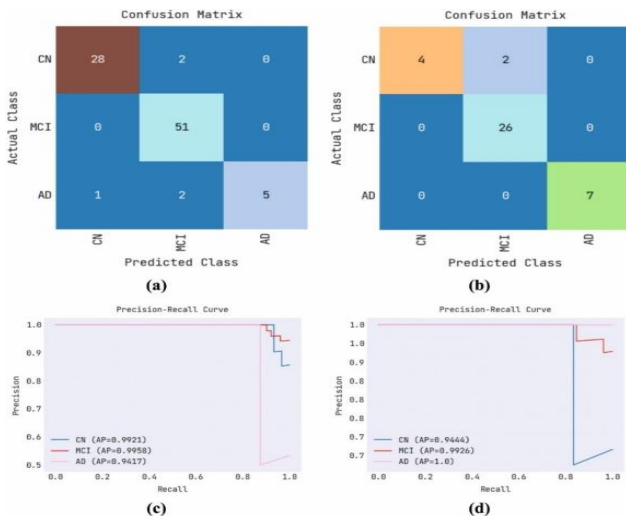


Fig. 4. (a) TR set confusion matrix (b) TS set confusion matrix (c) TR set PR-curve (d) TS set PR-curve.

Fig. 4 determines the classifier results of the WOADL-BDDC system under the training (TR) set and testing (TS) set. Fig. 4(a) represents the confusion matrix offered by the WOADL-BDDC model under the TR set. The figure symbolized that the WOADL-BDDC model had identified 28, 51, and 5 instances under CN, MCI, and AD classes. Also, Fig. 4(b) describes the confusion matrix offered by the WOADL-BDDC methodology under the TS set. The figure exemplified that the WOADL-BDDC algorithm has identified 4, 26, and 7 instances under CN, MCI, and AD classes. Fig. 4(c) and 4(d) validate the precision-recall analysis of the WOADL-BDDC system under training and TS sets. The figures described that the WOADL-BDDC method had acquired maximal precision-recall performance under all varying classes.

Table I reports the overall brain disorder classification outcomes of the WOADL-BDDC model. Fig. 5 represents a brief brain disorder classifier results of the WOADL-BDDC model under the TR set. The results indicated that the WOADL-BDDC model had shown enhanced results under all measures. It is noticed that the WOADL-BDDC model has gained an accuracy of 94.38%, percent of 96.43%, sense of 85.28%, species of 95.93%,  $F1_{score}$  of 89.35%, and FPR of 4.07%.

TABLE I. RESULT ANALYSIS OF THE WOADL-BDDC SYSTEM WITH DISTINCT MEASURES

Metrics	TR set	TS set
Accuracy	94.38	94.87
Precision	96.43	97.62
Sensitivity	85.28	88.89
Specificity	95.93	94.87
F1-Score	89.35	92.10
False Positive Rate	04.07	05.13

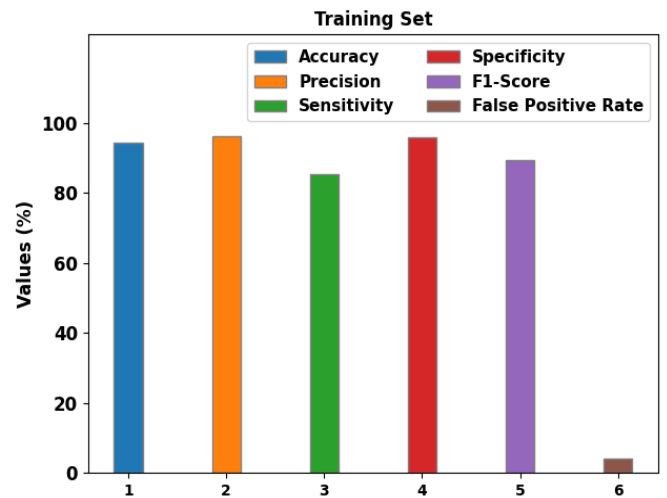


Fig. 5. Result in the analysis of the WOADL-BDDC system under the TR set.

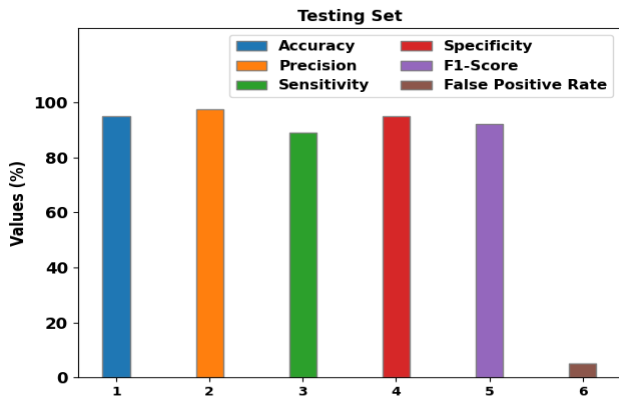


Fig. 6. Result in the analysis of the WOADL-BDDC system under the TS set.

Fig. 6 exemplifies a brief brain disorder classifier results of the WOADL-BDDC approach under the TS set. The outcomes referred that the WOADL-BDDC process has shown enhanced results under all measures. It is observed that the WOADL-BDDC system has gained an accuracy of 94.87%, percent of 97.62%, sense of 88.89%, species of 94.87%,  $F1_{score}$  of 92.10%, and FPR of 5.13%.

An obvious ROC analysis of the WOADL-BDDC approach on the TR set is shown in Fig. 7. The outcomes referred to the WOADL-BDDC algorithm have exhibited their ability to categorize distinct classes.

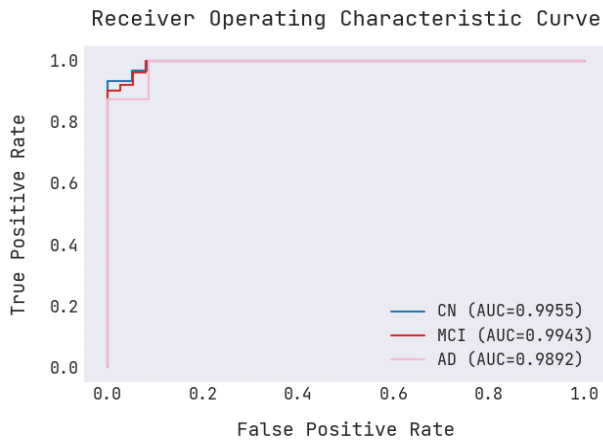


Fig. 7. ROC curve of WOADL-BDDC system under TR set.

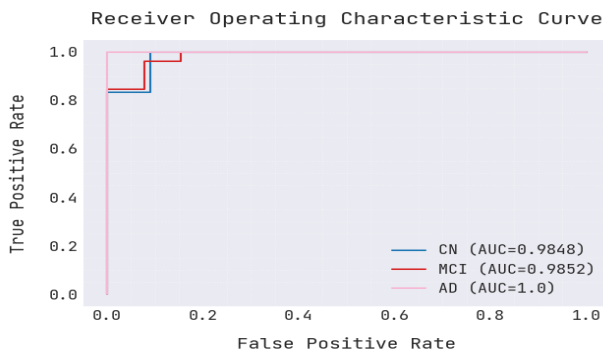


Fig. 8. ROC curve of WOADL-BDDC system under TS set.

A detailed ROC analysis of the WOADL-BDDC system on the TS set is displayed in Fig. 8. The outcomes stated the WOADL-BDDC approach had revealed its capability in varying classifier classes.

The training accuracy ( $TR_{acc}$ ) and validation accuracy ( $VL_{acc}$ ) acquired by the WOADL-BDDC system on the test database is depicted in Fig. 9. The simulation value stated that the WOADL-BDDC method had reached increased values of  $TR_{acc}$  and  $VL_{acc}$ . In particular, the  $VL_{acc}$  looked that greater than  $TR_{acc}$ .

The training loss ( $TR_{loss}$ ) and validation loss ( $VL_{loss}$ ) accomplished by the WOADL-BDDC system under the test database are displayed in Fig. 10. The simulation value pointed out that the WOADL-BDDC system has achieved lesser values of  $TR_{loss}$  and  $VL_{loss}$ . In particular, the  $VL_{loss}$  is lesser than  $TR_{loss}$ .

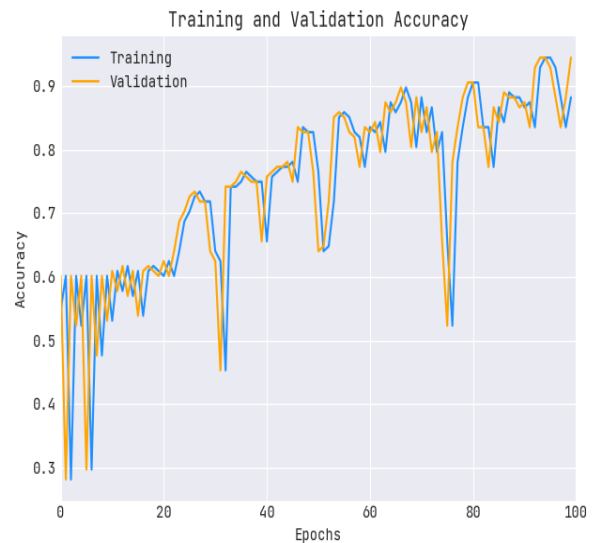


Fig. 9. Result analysis of WOADL-BDDC system in terms of  $TR_{acc}$  and  $VL_{acc}$ .

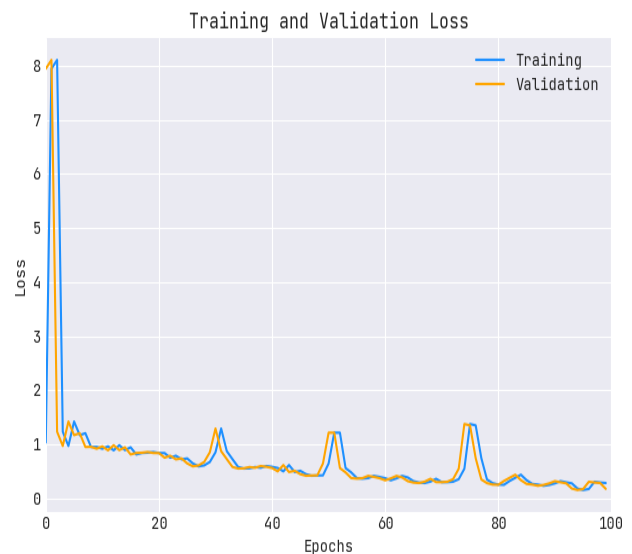


Fig. 10. Result Analysis of the WOADL-BDDC model in terms of  $TR_{loss}$  and  $VL_{loss}$ .

TABLE II. COMPARATIVE ANALYSIS OF THE WOADL-BDDC SYSTEM WITH RECENT SYSTEMS

Methods	Accuracy (%)	Specificity (%)
WOADL-BDDC	94.38	95.93
Vgg16	90.35	89.50
Graph CNN	94.00	93.02
Modified ResNet18	92.81	94.66
Non-linear SVM	75.58	75.63
ResNet50-SVM	92.49	87.36
ResNet50-RF	86.31	79.57

A comprehensive comparison study of the WOADL-BDDC model on brain disorder classification is shown in Table II [23, 24].

Fig. 11 represents a comparative investigation of the WOADL-BDDC technique with recent approaches in terms of accuracy. The results indicated that the non-linear SVM and ResNet50-RF models had attained the least accuracy values of 75.58% and 86.31%, respectively. Following, the VGG16 has shown a slightly enhanced accuracy of 90.35%. In contrast, the graph CNN, modified ResNet18, and ResNet50-SVM models have reached reasonable accuracy values of 94%, 92.81%, and 92.49%. Finally, the WOADL-BDDC model has obtained promising performance with a higher accuracy of 94.38%.

Fig. 12 signifies a comparative investigation of the WOADL-BDDC system with current methods in terms of species. The outcomes indicated that the non-linear SVM and ResNet50-RF approaches had correspondingly accomplished the least  $spec_y$  values of 75.63% and 79.57%. Next, the VGG16 has demonstrated a somewhat improved  $spec_y$  of 89.50%. In contrast, the graph CNN, modified ResNet18, and ResNet50-SVM models have reached reasonable  $spec_y$  values of 93.02, 94.66%, and 87.36%. Finally, the WOADL-BDDC methodology has attained promising performance with a higher  $spec_y$  of 95.93%.

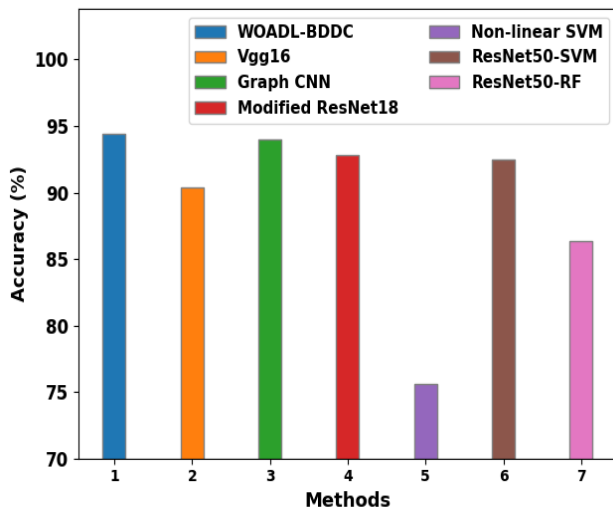


Fig. 11.  $Accu_y$  analysis of WOADL-BDDC system with recent algorithms.

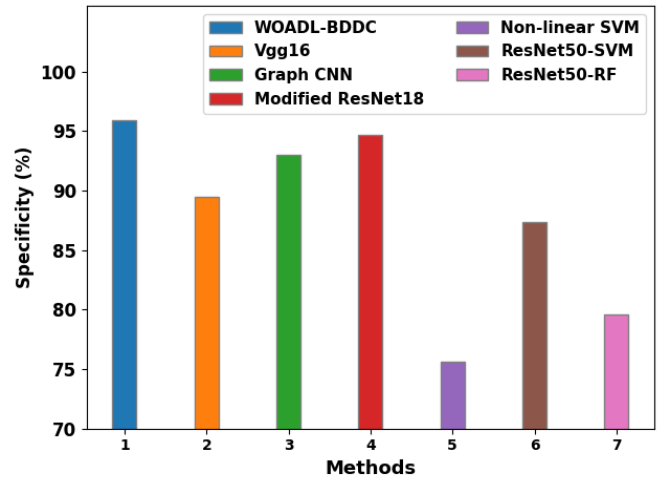


Fig. 12.  $Spec_y$  analysis of WOADL-BDDC system with recent algorithms.

These results confirmed the automated and accurate brain disorder classification results of the WOADL-BDDC model.

## V. CONCLUSION

In this study, we have established a novel WOADL-BDDC technique for recognizing brain disorders on MRI images. Primarily, the presented WOADL-BDDC technique performs pre-processing in two stages: GF-based noise elimination and U-Net segmentation-based skull stripping. Besides, the QuickNAT with RMSProp optimizer is applied to segment the brain parts. For feature extraction, the WOADL-BDDC technique performs a radiomics approach that separates features of all aspects. Finally, WOA with CRNN model is applied for brain disorder classification. The experimental evaluation of the WOADL-BDDC technique is tested on ADNI 3D dataset. The comprehensive comparison study highlighted the adequate performance of the WOADL-BDDC technique in different measures. In the future, an ensemble of DL-based classifiers can be employed to improve the WOADL-BDDC technique's detection performance.

## VI. FUNDING STATEMENT

The authors received no specific funding for this study.

## VII. CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest to report regarding the present study.

## REFERENCES

- [1] AlSaeed, D.; Omar, S.F. Brain MRI Analysis for Alzheimer's Disease Diagnosis Using CNN-Based Feature Extraction and Machine Learning. *Sensors*, 22, 2911, 2022. <https://doi.org/10.3390/s22082911>.
- [2] Altaf, T., Anwar, S.M., Gul, N., Majeed, M.N. and Majid, M., Multi-class Alzheimer's disease classification using image and clinical features. *Biomedical Signal Processing and Control*, 43, pp.64-74, 2018.
- [3] Alwajih, R., Abdulkadir, S.J., Al Hussian, H., Aziz, N., Al-Tashi, Q., Mirjalili, S. and Alqushaibi, A., Hybrid binary whale with harris hawks for feature selection. *Neural Computing and Applications*, pp.1-19, 2022.
- [4] An, N., Ding, H., Yang, J., Au, R. and Ang, T.F. Deep ensemble learning for Alzheimer's disease classification. *Journal of biomedical informatics*, 105, pp.103-112, 2020.

- [5] Basheera, S. and Ram, M.S.S., Convolution neural network-based Alzheimer's disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 5, pp.974-986, 2019.
- [6] Bi, X. and Wang, H., Early Alzheimer's disease diagnosis based on EEG spectral images using deep learning. *Neural Networks*, 114, pp.119-135, 2019.
- [7] Fouad, A., Mofteh, H. M., & Hefny, H. A., Brain diagnoses detection using whale optimization algorithm based on ensemble learning classifier. *Int. J. Intell. Eng. Syst*, 13(2), 40-51, 2020.
- [8] Hu, Y., Wen, C., Cao, G., Wang, J. and Feng, Y., Brain network connectivity feature extraction using deep learning for Alzheimer's disease classification. *Neuroscience Letters*, 782, p.136673, 2022.
- [9] Jo, T., Nho, K., Bice, P., Saykin, A.J. and Alzheimer's Disease Neuroimaging Initiative, Deep learning-based identification of genetic variants: application to Alzheimer's disease classification. *Briefings in Bioinformatics*, 23(2), p.bbca022, 2022.
- [10] Jo, T., Nho, K., Risacher, S.L. and Saykin, A.J., Deep learning detection of informative features in tau PET for Alzheimer's disease classification. *BMC bioinformatics*, 21(21), pp.1-13, 2020.
- [11] Arulananth, T.S., Balaji, L., Baskar, M., PCA Based Dimensional Data Reduction and Segmentation for DICOM Images. *Neural Process Lett* 2020. <https://doi.org/10.1007/s11063-020-10391-9>.
- [12] Khagi, B., Kwon, G.R. and Lama, R., Comparative analysis of Alzheimer's disease classification by CDR level using CNN, feature selection, and machine-learning techniques. *International Journal of Imaging Systems and Technology*, 29(3), pp.297-310, 2019.
- [13] Khan, M.A., HCRNNIDS: hybrid convolutional recurrent neural network-based network intrusion detection system. *Processes*, 9(5), p.834, 2021.
- [14] Li, L. and Ma, H., Pulse coupled neural network-based multimodal medical image fusion via guided filtering and WSEML in NSCT domain. *Entropy*, 23(5), p.591, 2021.
- [15] Liang, G., Xing, X., Liu, L., Zhang, Y., Ying, Q., Lin, A.L. and Jacobs, N., Alzheimer's disease classification using 2d convolutional neural networks. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 3008-3012). IEEE, 2021.
- [16] Odusami, M.; Maskeliunas, R.; Damasevicius, R.; Krilavicius, T. Analysis of Features of Alzheimer's Disease: Detection of Early Stage from Functional Brain Changes in Magnetic Resonance Images Using a Finetuned ResNet18 Network. *Diagnostics* 2021, 11, 1071. <https://doi.org/10.3390/diagnostics11061071>.
- [17] Qiu, S., Joshi, P.S., Miller, M.I., Xue, C., Zhou, X., Karjadi, C., Chang, G.H., Joshi, A.S., Dwyer, B., Zhu, S. and Kaku, M., Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. *Brain*, 143(6), pp.1920-1933, 2020.
- [18] Roy, A.G., Conjeti, S., Navab, N., Wachinger, C. and Alzheimer's Disease Neuroimaging Initiative, QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy. *NeuroImage*, 186, pp.713-727, 2019.
- [19] Sampath, R. and Baskar, M., 3D brain image-based Alzheimer's disease detection techniques using fish swarm optimizer's deep convolution Siamese neural network. *Expert Systems*, p.e12963, 2022.
- [20] Shanmugam, J.V., Duraisamy, B., Simon, B.C. and Bhaskaran, P., Alzheimer's disease classification using pre-trained deep networks. *Biomedical Signal Processing and Control*, 71, p.103217, 2022.
- [21] Sharma, R., Goel, T., Tanveer, M. and Murugan, R., FDN-ADNet: Fuzzy LS-TWSVM based deep learning network for prognosis of the Alzheimer's disease using the sagittal plane of MRI scans. *Applied Soft Computing*, 115, p.108099, 2022.
- [22] Tian, J., Smith, G., Guo, H., Liu, B., Pan, Z., Wang, Z., Xiong, S. and Fang, R. Modular machine learning for Alzheimer's disease classification from retinal vasculature. *Scientific Reports*, 11(1), pp.1-11, 2021.
- [23] Wang, Y., Liu, W., Yu, Y., Liu, J.J., Xue, H.D., Qi, Y.F., Lei, J., Yu, J.C. and Jin, Z.Y., CT radiomics nomogram for the preoperative prediction of lymph node metastasis in gastric cancer. *European radiology*, 30(2), pp.976-986, 2020.
- [24] Zhang, J., Zheng, B., Gao, A., Feng, X., Liang, D. and Long, X., A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification. *Magnetic Resonance Imaging*, 78, pp.119-126, 2021.
- [25] Amin, J., Sharif, M., Haldorai, A., Yasmin, M., and Nayak, R. S., Brain tumor detection and classification using machine learning: a comprehensive survey. *Complex Intell. Syst.* 8, pp.3161-3183, 2021. doi: 10.1007/s40747-021-00563-y.
- [26] Baskar, M., Renuka Devi, R., Ramkumar, J., Region Centric Minutiae Propagation Measure Orient Forgery Detection with Finger Print Analysis in Health Care Systems. *Neural Process Lett*. 2021. <https://doi.org/10.1007/s11063-020-10407-4>.
- [27] Houssein, E. H., Ibrahim, I. E., Neggaz, N., Hassaballah, M., and Wazery, Y. M., An efficient ecg arrhythmia classification method based on manta ray foraging optimization. *Expert. Syst. Appl.* 181, 115131, 2021. doi: 10.1016/j.eswa.2021.115131.
- [28] Bekhet, S., Hassaballah, M., Kenk, M. A., and Hameed, M. A., An artificial intelligence based technique for COVID-19 diagnosis from chest x-ray, in *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)* (Giza: IEEE), pp.191-195, 2020.

# A Study on the Designation Institution for Supercomputer Specialized Centers in Republic of Korea

Hyungwook Shim, Yonghwan Jung, Jaegyoon Hahm\*

National Supercomputer Department, Korea Institute of Science and Technology Information, Daejeon, Republic of Korea

**Abstract**—In Korea, specialized centers are designated for 10 strategic fields for the purpose of jointly utilizing supercomputer resources at the national level. Based on the “National Supercomputing Innovation Strategy,” it plans to select 10 centers in three stages by 2030, and has now completed the designation of the first-stage specialized centers in 2022. With the second designation in 2024 ahead, it is urgent to review and improve the existing designation institution for fairer and more effective selection of specialized centers. Therefore, this paper analyzed the influence of evaluation items and the influence of evaluation items on evaluation results by using logistic regression analysis and network centrality analysis to prepare improvement plans for the existing evaluation model. As a result of the analysis, improvement measures were derived, such as subdividing evaluation items with low impact, expanding the items, and lowering the allotment of evaluation items with low impact.

**Keywords**—Supercomputer; specialized center; evaluation system; logistic regression model; network centrality analysis

## I. INTRODUCTION

Korea's supercomputer governance consists of a national center, a specialized center, and a unit center. The national center secures and operates supercomputing resources, supports policy establishment, and manages joint utilization. The specialized center performs supercomputing resource establishment and operation, basic application research and dissemination of results, etc. The unit center is a resource independently operated by individual private research institutes and companies [1]. Specialized centers maintain their qualifications for five years after designation. Currently, designation of specialized centers in 7 fields has been completed, and by 2028, it will be expanded to 10 fields[2]. Recently, the Ministry of Science and ICT announced that it would establish the “3rd supercomputer development basic plan”(referred to as '3rd Basic Plan'), the top plan for supercomputers, and establish a user support system centered on specialized centers. Therefore, at the beginning of the "3rd Basic Plan", the government should make efforts to improve the special center designation institution to ensure fairness, effectiveness, and sustainability. As a measure to improve the government's evaluation system in the field of science and technology, a statistical method using the influence of each evaluation item on the evaluation result is widely used [3]. Therefore, this paper also presents a plan to improve the specialized center designation institution by using the

evaluation results for the designation of the existing specialized center.

This paper consists of six sections. Sections I and II presented an academic value through a qualitative analysis of the background, meaning, and source research of this paper. Section III introduces the function, role, and protection system of the supercomputer specialized center, and Section IV explains the methodology of this paper. In Section V, a case study for improving the evaluation model is conducted and the results are gradually presented. Finally, in Section VI, the results were summarized and the viewpoint was straightened out, and the final point and pursuit plan of this paper were presented.

## II. LITERATURE REVIEW

Major prior studies are as follows. Hirao (2010) introduced projects for the introduction of peta-class next-generation supercomputing systems [4], and Hsu (2015) analyzed foreign trends for exascale supercomputing development and introduced major projects invested in the United States [5]. Mitsuhsa (2021) introduced Fugaku's flagship project related to Japan's Fugaku supercomputer and presented design details such as Fugaku's scale and performance [6]. Savin (2019) introduced the supercomputing center community system in Russia and mentioned the advantages in terms of energy efficiency and the provision, monitoring, and management of resources through a shared utilization network. In addition, improvement plans were presented through analysis of the current status of the Joint Supercomputer Center [7]. Prior domestic studies are as follows. Huh (2021) conducted research on ways to improve the legal system to vitalize the supercomputing ecosystem in Korea. Regarding the supercomputer-related law, the ‘Supercomputer Act’, problems such as the role of related institutions, project costs, mutual cooperation system, and consistency with higher-level plans were identified, and improvement measures were proposed with a focus on policy consistency and effectiveness enhancement [8]. Shim (2022) conducted a study to improve the evaluation index for selecting a research institute for the national R&D project of the Ministry of Land, Infrastructure and Transport. Using the evaluation score of each evaluator, the evaluation index was determined through an artificial neural network, and a method for improving the score distribution for each evaluation index was derived using logistic regression analysis [9]. Shin (2013) conducted a study to prepare improvement plans for local government

\*Corresponding Author.

performance indicators, classified performance indicators by characteristics, and applied development procedures that considered visions and promotion strategies for each local government and key outcomes by function. In addition, improvement measures were derived considering desirable performance indicator attributes such as relevance, clarity, timeliness, reliability, and comparability [10]. Lee (2018) set the field and elements for improving the educational environment and set the direction of the indicators through a delphi survey by experts for the purpose of developing educational environment improvement indicators. As a result of the analysis, the indicators for improving the educational environment, such as the adequacy of the total floor area of the classroom, the adequacy of general school teaching, whether or not to secure seismic performance, the deterioration of firefighting facilities, and energy consumption, were finally determined [11]. Ji (1999) conducted research for the rational development of informatization indicators that measure the national informatization level. The author first set informatization facility indicators, informatization use indicators, and informatization support indicators, subdivided them into 6 groups, and proposed an informatization measurement indicator system consisting of a total of 28 indicators in consideration of informatization level, reality, and applicability aspects[12]. Kim (2022) conducted a study on how to improve the global cyber security index, which is used to diagnose the level of national cyber security development and strengthen cyber security capabilities. The author established basic principles for the improvement and utilization of the Global Cyber Security Index and suggested development plans through survey-based SWOT analysis[13].

The academic value of this paper is as follows. It is novel because no research has been conducted on the improvement of evaluation system related to the existing domestic supercomputer. Although Huh (2021) conducted a study on institutional improvement measures related to supercomputers, this paper only examines the appropriateness of the evaluation system. Also, research related to the development and improvement of existing evaluation system draws conclusions using a qualitative method of asking and organizing the opinions of experts. However, in this paper, the influence of each index of the existing evaluation model was analyzed using a statistical method using the actual evaluation result data of the evaluation committee, and a comparative analysis was performed with the improved evaluation model. Lastly, a survey was conducted on the appropriateness of the evaluation index targeting the researchers of the specialized center support institution that was evaluated, and the fairness of the improvement model was added by reflecting the opinions of all parties participating in the evaluation.

### III. SUPERCOMPUTER SPECIALIZED CENTER

#### A. Definition of Supercomputer Specialist Center

Supercomputer specialized center defined as an institution that possesses expertise for professional use of supercomputers, provides specialized services based on resources, manpower, and technology specialized in the field, conducts research and development, and promotes the use of supercomputers. The functions and roles of the specialized

centers in Table I include the establishment and operation of supercomputing resources by field, service provision, base application research and dissemination of research results, large-capacity data management and operation support, and human resource training.

#### B. Evaluation Institution for Designation of Supercomputer Specialized Center

The designation of specialized centers is in accordance with the "Operational Guidelines for Designating Supercomputers by Field" (referred to as 'Operational guidelines'). The Operating guidelines are based on Article 9-2 of the "Act on the Promotion and Utilization of Supercomputing" (referred to as the 'Supercomputer Act'), which includes the functions and roles of specialized centers, designation procedures and methods, establishment of operation plans, evaluation of operational performance, composition and operation of evaluation teams, etc. The designation procedure is shown in Fig. 1. First of all, the institutes, universities target organizations that can submit applications are central administrative agencies, national and public research, private companies and organizations with expertise in each field, and meet the requirements such as 4 or more supercomputer experts and supercomputer possession (1.5 million dollars or more). After submitting the application, one institution in each field can be designated as a specialized center through the first written examination, the holding of a briefing session, and the second face-to-face examination. Finally, among institutions with a score of 70 or more, the institution with the highest score is selected.

In the evaluation system, the subject of evaluation is the evaluation team, and it is made up of 3 or more and 10 private experts, including the head. The evaluation team conducts both the first and second evaluation, and the main evaluation items are shown in Table II. The items and indicators were derived through the FGI of experts in the related field, and the evaluation items include the 'Related Performance', 'Validity of operation purpose and plan', 'Suitability of center manpower', etc.

TABLE I. FUNCTIONS AND ROLES OF THE SPECIALIZED CENTERS

	Main Content
Function and Role	<ul style="list-style-type: none"> <li>· Providing supercomputing services in specialized fields</li> <li>· research and development of supercomputing technology and application specialized in specialized fields</li> <li>· xpansion of resource sharing through participation in the joint utilization system</li> </ul>
Service	<ul style="list-style-type: none"> <li>· stablishment, operation and service of specialized resources for each sector centered on 10 strategic fields</li> <li>· tilization support and consulting through field experts</li> <li>· Community development and education by field</li> </ul>

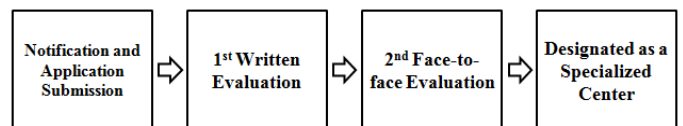


Fig. 1. Specialized center designation procedure.



TABLE II. EVALUATION ITEM AND INDICATOR

Evaluation item	Evaluation indicator	points
Related Performance	Performance of R&D manpower training	10
	Performance of service, data management and research support	10
Validity of operation purpose and plan	Justification and Necessity of Designation	5
	Challenge and specificity of vision and operational goals	10
	Suitability of goals, project contents, research methods, etc.	10
	Excellence of Expected Performance and Utilization Plan	5
	Center fostering and operation support plan	10
Suitability of center manpower	Appropriateness of personnel planning and organizational composition	10
	Researcher's expertise	10
Facility and Equipment Securing Plan	Facility, Equipment Status and Expansion Plan	15
	Facility Status and Expansion Plan	5
Added points	Ratio of joint utilization resources	10

In the specialized center designation institution, the evaluation items and indicators for designation are important factors in selecting a specialized center representing a specific field for the next five years. Regarding the national R&D project evaluation system, Lee (2010) also emphasizes the effects of the need for improvement of evaluation items and indicators, and the preparation of improvement plans using evaluation results [14]. Using the actual evaluation result data, the appropriateness of new discovery of evaluation items, removal of existing items, and adjustment of points allocation was proved. Therefore, in this paper, it is necessary to review the appropriateness of evaluation items based on the evaluation results of the seven specialized centers.

#### IV. THEORETICAL BACKGROUND AND METHODOLOGY

##### A. Research Procedure

In many papers such as Ahn(2022), regression analysis, AHP, and machine learning analysis are mainly used to research measures to improve the evaluation system [15]. The analysis results are effective in discovering individual improvement factors for evaluation items, and can provide intuitive results to researchers. However, evaluation items are grouped into various sub-indicators, and due to the nature of R&D, the range of evaluation items such as research method, content, and research timing is radially intertwined. Therefore, it is more effective to discover improvement factors by considering the correlation between evaluation items, and in this paper, an improvement plan was derived considering the result of network centrality analysis to reflect the correlation between evaluation items. The research procedure of this paper is shown in Fig. 2. First, using logistic regression analysis for the existing evaluation model, the influence on the selection result for each evaluation item and the appropriateness of the points assigned are reviewed. Second, through network centrality analysis, the evaluation model improvement plan is derived by analyzing the structure and centrality that affect evaluation items. Finally, the results are summarized and implications are drawn.

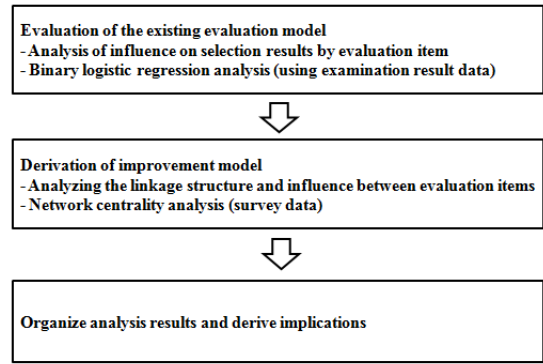


Fig. 2. Specialized center designation procedure.

##### B. Logistic Regression Model Analysis

Binary logistic regression analysis is used to estimate the relationship between a binary dependent variable (categorical) and multiple independent variables to explain the influence of the independent variable on the dependent variable or to predict the value of the dependent variable for the value of the independent variable. It is an analysis technique that regression models can generally be expressed as in Equation (1), and  $\hat{y}$  has values of 0 and 1.

$$\hat{y} = b_0 + b_1X_1 + b_2X_2 \cdots b_kX_k + \epsilon \quad (1)$$

As shown in Fig. 3, the regression model can be divided into total variance SST (Total Sum of Squares), which means the difference between the actual value and the mean, and error variance SSE (Error Sum of Squares), which means the difference between the actual value and the estimated value, and variance by the regression equation SSR (Sum of Squares due to Regression). Through these three fluctuation values, the coefficient of determination, which means the contribution to explain the diversity of the dependent variable, is obtained, and the regression coefficient  $b$  can be estimated as shown in Equation (2) through the least squares method in which the variation in error(SSE) is minimized.

$$b_k = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_{ij} - \bar{x}_i)^2} \quad (2)$$

In this paper, the selection result was substituted as a dependent variable, and '0' was set as not selected and '1' as selected. Independent variables represent 4 evaluation items excluding added points, and were analyzed by standardizing them to values between 0 and 1.

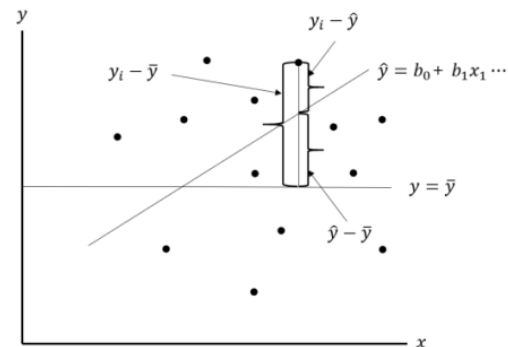


Fig. 3. Regression model conceptual diagram.

C. Network Centrality Analysis

Network analysis is an analysis method that analyzes the characteristics of relationships that exist between objects (nodes). It can visualize a microscopic network of relationships by deriving characteristics such as connection strength and connection structure between research subjects using nodes and links. Network centrality analysis is an analysis method that utilizes centrality indicators using the number of connections between nodes, distances, and travel routes among various measurement indicators [16]. Connectivity centrality, proximity centrality, betweenness centrality, and eigenvector centrality are typically used as centrality indicators [17]. The eigenvector centrality index used in this paper is a method of analyzing centrality by weighting the centrality of the other node. Characteristically, as shown in Equation (3), the maximum eigenvalue  $\lambda$  of the matrix between nodes is used.

$$C_i = \frac{1}{\lambda} \sum_j \alpha_{i,j} C_j \quad (3)$$

$i, j$ : Node

$C$ : Eigenvector centrality

$\lambda$ : eigenvalue

$\alpha$ : Connection relationship (connected 1, unconnected 0)

Nodes mean evaluation items during analysis, and using the matrix constructed using the survey results, eigenvalues are calculated and eigenvector centrality for evaluation items and indicators is estimated [18].

V. CASE STUDY

A. Data

Data for case studies can be classified into two types. First, the data for the logistic regression analysis use the evaluation result data of the specialized center evaluation team. The evaluation result data uses written evaluation data for 7 fields, and uses the evaluation scores written by the evaluation team for a total of 15 institutions in 7 fields. The dataset is shown in Table III. A, B, C, and D were selected as independent variables and entered as 0~1 scale. The dependent variable was selected as a nominal variable (dummy variable) with two values of 0 and 1 depending on whether or not it was selected.

Next, as data for network centrality analysis, data from an online survey targeting 60 people, including executives and employees belonging to seven specialized centers, are used. The survey items consist of a total of 17 items, including a total of 16 items in the Likert scale method and one item in the matrix method.

B. Logistic Regression Analysis Results

The analysis results of the regression model including all evaluation items are shown in Tables IV and V. Table IV shows the Nagelkerke R2 index, which indicates the explanatory power of the entire model, and the Hosmer-Lemeshow index, which is a goodness-of-fit test of the model. In the Hosmer-Lemeshow test, the chi-square value indicates the degree of agreement between the actual value of the dependent variable and the predicted value by the model and

the smaller the chi-square value, the higher the fitness of the model. The extent to which independent variables explained the dependent variable was about 24%. Since the value of the significance probability  $p$  in the goodness of fit of the model was larger than 0.05, the null hypothesis was not rejected, so the goodness of fit can be considered acceptable.

Table V is the estimation result of the regression coefficient of the model. First, if the sign of the regression coefficient  $\beta$  is positive (+), the greater the value of the corresponding independent variable, the greater the possibility of being classified as a selected group representing the dependent variable '1', and negative (-) means the opposite case. As a result of the analysis, evaluation items A and D are positive (+), and the higher the score of evaluation items A and D, the higher the possibility of being selected. B and C are negative (-), and the higher the score, the higher the possibility of not being selected. The significance probability is less than 0.05 for both evaluation items A and D in the 95% confidence interval, which can be considered significant, and B and C are 0.227 and 0.901, respectively, which are greater than 0.05, so it can be considered insignificant. Wald is a statistic that verifies whether the coefficient value for each covariate is zero.  $\text{Exp}(\beta)$  represents the odds ratio and means the influence on the evaluation result when an evaluation item increases by one unit. It can be interpreted that when the score of evaluation item A increases by one unit, the probability of being selected increases about 20 times.

TABLE III. DATASET

Item	Indicator	Definition	Data coding
Independent variable	A	Related Performance	0~1 (Scale)
	B	Validity of operation purpose and plan	
	C	Suitability of center manpower	
	D	Facility and Equipment Securing Plan	
Dependent variable	Result	-	0, 1 (Dummy)

TABLE IV. ANALYSIS RESULT OF HOSMER-LEMESHOW TEST

	Chi-square	Significance probability	Note
Hosmer-Lemeshowtest	8.120	0.422	Nagelkerke R2 :.240

TABLE V. ESTIMATION RESULT OF THE REGRESSION COEFFICIENT

	B	S.E.	Wald	Significance probability	Exp(B)
A	3.014	1.449	4.328	.037	20.378
B	-.846	.700	1.462	.227	.429
C	-.243	1.958	.015	.901	.785
D	2.698	1.162	5.393	.020	14.849
Constant	-5.741	1.770	10.520	.001	.003

As a result of the statistical analysis of Table VI, it can be confirmed that the evaluation items B and C are not appropriate in the direction and significance probability affecting the evaluation result. In the case of the significance probability, it cannot be a factor that absolutely determines the validity of the independent variable, but in the case of the direction, it can be a factor that can determine the validity of the independent variable in consideration of the evaluation criteria. Therefore, after excluding the two evaluation items with a negative (-) sign, the regression coefficient was re-estimated, and the re-estimation results are shown in Table VII. The regression coefficients of the A and B evaluation items showed positive (+) values, and it was confirmed that they were somewhat reduced compared to the previous ones. The significance probability was statistically significant at the 90% confidence interval, and it was confirmed that the influence of the evaluation result was somewhat lowered according to the odds ratio result. The influence of D on the selection result was about 1.6 times greater than that of A. Through this, it is necessary to improve evaluation items B and C as a method for modifying the model. In order to prepare improvement measures, the network centrality analysis results are additionally conducted and the two analysis results are comprehensively considered.

TABLE VI. RESULT OF THE REGRESSION COEFFICIENT ESTIMATION

	B	S.E.	Wald	Significance Probability	Exp(B)
A	1.638	.918	3.188	.074	5.147
D	2.102	1.080	3.788	.052	8.184
Constant	-6.297	1.780	12.515	.000	.002

### C. Network Centrality Analysis Result

The result of visualizing the network for evaluation items is shown in Fig. 4. Node A represents relevant performance, node B represents the validity of operation purpose and plan, node C represents the suitability of center manpower, node D represents the facility and equipment securing plan, and node E represents the evaluation items for the add points. The size of a node increases as the frequency of the node increases, and the frequency is determined by the number of choices made by the respondent. The link is expressed as a straight line connecting the nodes, and the higher the co-occurrence frequency, the bolder it is.

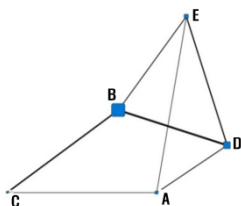


Fig. 4. Regression model conceptual diagram.

As a result of the analysis, the frequency of nodes was highest in evaluation item B, which means the validity of the operation purpose and plan, followed by D, E, A, and C in order. As for the links, the B-D link that connects the validity of the operation purpose and plan and the add points evaluation

items appeared in the thickest form, followed by B-C and D-E in that order. In other words, it can be analyzed that the strength of the relationship between B-D, B-C, and D-E is relatively strong and the strength of the relationship between the evaluation items C-A and A-E is relatively weak around evaluation item B in the network between evaluation items evaluated by the response group. The strength of the relationship can be interpreted in various ways, but from the perspective of the evaluation system, it can be interpreted in two cases. First, the allocation of evaluation items with high relationship strength should be relatively higher than those with low relationship strength. This is because the higher the relationship strength, the greater the effect on the overall evaluation scores. Second, it is necessary to distinguish between evaluation item groups with high relationship strength and evaluation item groups with low relationship strength. If the impact on the evaluation results is significantly large due to the difference in relationship strength, it may be appropriate to classify or exclude groups of evaluation items. Therefore, in this paper, when considering the improvement of the overall evaluation system, the relationship strength was used as a basis for adjusting the evaluation item. In order to analyze the quantitative influence of each evaluation item in the network, the centrality value is used, and the eigenvector centrality value, which considers the centrality value of other related evaluation items among the centrality values, was estimated. The results of the eigenvector centrality analysis for the evaluation items are shown in Table VII. The evaluation item with the greatest eigenvector centrality is B (0.572), followed by D (0.550), E (0.393), C (0.355), and A (0.300) in that order appears. In the quantitative analysis results, the centrality values of evaluation items B and D were relatively high, and E, C, and A were low.

TABLE VII. EIGENVECTOR CENTRALITY

	A	B	C	D	E
Eigenvector centrality value	0.300	0.572	0.355	0.550	0.393

Based on the analysis results, it can be divided into two groups according to the size of the centrality value. It can be divided into two groups: B, D, E, C, and A. When improving the evaluation system, it is appropriate to increase the score for the group with high relationship strength and to adjust the score for the group with low relationship strength.

The comprehensive improvement plan for the evaluation model, including the result of network centrality analysis, was determined by segmentation of evaluation item B. The reason for this decision is: First, it was derived as the most influential indicator in the model as a result of network centrality analysis, while being an object that needs improvement according to the results of regression analysis. Second, in the case of evaluation items B and D, the centrality value is similar, but the score of B is twice as high, so it is necessary to adjust the score of evaluation item B downward. Third, evaluation item B has the largest number of sub-indicators, so there is a limit to representing all the characteristics of sub-indicators. Fourth, in the case of improvement evaluation item C according to the results of regression analysis, it is appropriate to improve B evaluation item first because it has a low influence on other

evaluation items and is clearly classified as a group with B evaluation item. Therefore, this paper subdivided evaluation item B into two evaluation items. B1 is 'Challenge and specificity of vision and operational goals', 'Excellence of Expected Performance and Utilization Plan', B2 is 'Justification and Necessity of Designation', 'Suitability of goals, project contents, research methods, etc.', 'Center It was grouped under 'fostering and operation support plan'. For the improved model, logistic regression analysis is re-executed to examine the validity of the improvement.

#### D. Improvement Model Evaluation

The analysis results for the improvement model are shown in Tables VIII and IX. Through the Nagelkerke R2 index and the Hosmer-Lemeshow, a goodness-of-fit test of the model, the degree of explanation of the dependent variable by independent variables was about 33.6%, which was about 12% improved. As for the goodness of fit of the model, the value of the significance probability p was greater than 0.05, so the improved model also did not reject the null hypothesis.

TABLE VIII. ANALYSIS RESULT OF HOSMER-LEMESHOWTEST

	Chi square	Significance Probability	Note
Hosmer-LemeshowTest	5.772	0.673	Nagelkerke R2 :.336

Table IX is the estimation result of the regression coefficient of the improvement model. As a result of the analysis, all evaluation items are positive (+), and the higher the score, the higher the possibility of being selected. As for the level of significance, A, C, and D evaluation items were found to be significant at 95% confidence interval. B1 and B2 were found to be insignificant, but it was confirmed that they were greatly improved compared to the existing model. In addition,  $Exp(\beta)$  was similar for all evaluation items within 1.349 to 1.466, and the influence of B1 and B2 also increased by three major levels compared to the value of 0.429 in the existing model.

TABLE IX. ESTIMATION RESULT OF THE REGRESSION COEFFICIENT

	B	S.E.	Wald	Significance Probability	Exp(B)
A	.369	.189	3.814	.050	1.446
B1	.365	.246	2.201	.138	1.440
B2	.359	.195	3.381	.066	1.432
C	.383	.128	8.866	.003	1.466
D	.300	.104	8.234	.004	1.349
Constant	-22.980	6.606	12.102	.001	.000

### VI. CONCLUSION

This paper identified problems in which the initial evaluation items and points assigned in the national supercomputer specialized center designation evaluation institution were consistently applied until the end of the project, and proposed a sustainable evaluation model improvement plan using evaluation result data. The difference from previous studies is that a network centrality analysis was

newly performed to quantitatively analyze the strength of the relationship between evaluation items, and it was reflected in the improvement of evaluation items and indicators. As a result of the analysis, it was confirmed that the improvement plan of regrouping and subdividing the evaluation items using the eigenvector for the evaluation items was appropriate. It is expected that the results of this thesis will be used to continuously improve the designation institution and that excellent specialized centers will be selected.

The limitations of this thesis are that it has not been able to secure a lot of evaluation result data because many specialized centers have not yet been designated, and the effect of the improvement model has been proven only through statistical analysis. Therefore, in the 2023 second specialized center designation stage, plans are being established to apply the improved model in consultation with government agencies and re-verify the effect using the evaluation result data.

#### ACKNOWLEDGMENT

This research was carried out through major project (K-23-L02-C03-S01) support of Korea Institute of Science and Technology Information.

#### REFERENCES

- [1] "National Supercomputer Innovation Strategy", Republic of Korea, 2021.
- [2] "Supercomputer Specialized Center Operation Guidelines", Republic of Korea, 2021.
- [3] H.W. Shim, "An Analysis of Influence on the Selection of R&D Project by Evaluation Index for National Land Transport R&D Project" Journal of industrial convergence, Republic of Korea, vol. 20(2), pp.1-9, 2022.
- [4] K. Hirao, "The Next-Generation Supercomputer Project and a Plan for the Advanced Institute for Computational Science"Conference Proceedings of the international conference on supercomputing, vol. 24, pp. 3-4, 2010.
- [5] J. Hsu, "When will we have an exascale supercomputer?" IEEE spectrum, vol. 52(1), pp. 13-16, 2015.
- [6] S.Mitsuhsa, K. Yuetsu, T. Miwako, O. Tesuya, "Co-Design and System for the Supercomputer "Fugaku"" IEEE micro, vol. 42(2), pp. 26-34, 2021.
- [7] G. I. Savin, B.M. Shabanov, P. N.Telegin, A. V. Baranov, "Joint Supercomputer Center of the Russian Academy of Sciences: Present and Future", LOBACHEVSKII JOURNAL OF MATHEMATICS, vol. 40(11), pp. 1853-1862, 2019.
- [8] T.S.Huh, Y.H.Jung, M.J.Koh,"The Journal of the Korea Contents Association", Legal Institutional Improvement for Activating National Supercomputing Ecosystem, vol. 21(2), pp. 641-651, 2021.
- [9] H.W. Shim, "An Analysis of Influence on the Selection of R&D Project by Evaluation Index for National Land Transport R&D Project" Journal of industrial convergence, vol. 20(2), pp. 1-9, 2022.
- [10] H.C. Shin, "A Study on the Use of Performance Measures by Korean Local Governments: Limitations and Improvement Possibilities" Journal of Business Research, vol. 28(1), pp. 25-56, 2013.
- [11] S.M. Lee,H.J. Choi, " A Study on the Use of Performance Measures by Korean Local Governments: Limitations and Improvement Possibilities".
- [12] The journal of Sustainable Design and Educational Environment Research, vol. 17(3), pp. 35-45, 2018.
- [13] K.Y.Jee, S.W.Kang, J.R.Kim, "A Study for Refining the National Information Indicators)", Electronics and telecommunications trends, vol. 14(2), pp. 14-25, 1999.
- [14] D.K.Kim, J.H.Lee, Y.Y.Kim, D.E. Hyeon, H.R.Oh, "Proposals for GCI Indicators to Improve a National Cybersecurity Level", vol. 32(3), pp. 14-25, 2022.

- [15] H.J.Lee, "An Ontological Approach to Select R&D Evaluation Metrics" *Journal of the society of Korea industrial and systems engineering*, vol. 33(1), pp. 80-90, 2010.
- [16] S.Y.Ahn, Y.H. Cho, "A Study on the Performance Indicators for Evaluating the Qualitative Performance of National R&D Programs" *Journal of Korea Technology Innovation Society*, vol. 30(3), pp. 35-64, 2022.
- [17] H.W. Shim, "A Study on Improvement of Seoul Bike Sharing Service Usage Rate based on Network Centrality Analysis", *journal of Korean Society of Transportation*, vol. 37(2), pp. 124-134, 2019.
- [18] K.H.Lee, "A Study on the Changes in Maritime Network Centrality of the Busan Port : Focusing on Transpacific Route and Asia-Europe Route" *Journal of Shipping and Logistics*, vol. 37(3), pp 689-714, 2021.

# Time Series Forecasting using LSTM and ARIMA

Khulood Albeladi, Bassam Zafar, Ahmed Mueen

Department of Information Systems-Faculty of Computing & Information Technology, King Abdulaziz University, Jeddah, KSA

**Abstract**—Time series analysis is the process of evaluating sequential data to extract meaningful statistics. In the current era, organizations rely greatly on data analysis to solve and predict possible answers to a specific problem. These predictions help greatly in decision-making. In time series problems, the data is used to train the different machine and deep learning models. The models train on provided data displays particular outcomes. These outcomes anticipate possible solutions. In this paper, the two most effective Python models LSTM (Long Short Term Memory Loss) and ARIMA (Autoregressive integrated moving average) are used. These are the two most recommended models while dealing with Time series forecasting. The selected dataset is from Mulkia Gulf Real Estate available at MarketWatch. The main objective of this research paper is to study and compare the results of the two models used and determine which one is the best-suited model for that particular type of prediction. However, these are widely used models but the focus point of this research is determining the performance variance between these two models. LSTM became famous in 1997 as a training model that can remember patterns based on previous data while ARIMA is famous for forecasting variables of interest using a linear combination of previous values of the variable. The findings state that ARIMA is better for time series forecasting than LSTM based on the mean average of the basic evaluation parameters.

**Keywords**—Stock analysis; machine learning; deep learning; time series; ARIMA; LSTM

## I. INTRODUCTION

Time series is a time-dependent dataset, which means that the values are obtained in specific intervals of time. Usually, the values are taken at regular intervals, but the sampling could be irregular [4]. If a time series has a definite pattern, then any value of the series should be a function of previous values. Time series models differ from others in the way it predicts. With the advancement of information technology, now there are many more ways to collect time series data. A time series model uses a lag value of the target variable and uses it as a predictor variable, whereas traditional models use other variables as predictors. Time series analysis is the process of extracting output from time series data using different techniques. One of the famous types of time series analysis is Time series forecasting. In time series forecasting, the results are the predicted outputs from the trained models. There are many forecasting models available. In this research, LSTM and ARIMA models are used.

A general LSTM model has a cell; each cell has three parts, i.e. Forget gate, Input Gate and Output Gate. The states are affected by both past states and current input with feedback connections. LSTM models are able to learn long-term dependencies with feedback connections. LSTM can be single or multilayered models, the functionality differs for

both. They are preferred because they hold information for a long time by default. ARIMA forecasts temporal dependencies using only historical values. These models help to gain better insights into the data and predict future trends. It works by stationarising the series, means studying the correlation of the values and check its residual diagnosis ACF and PACF plots.

A number of forecasting problems can be solved using LSTM and ARIMA models. Although, the models have been used for prediction in the past (21) 19 18 12, but the goal of this research is to elaborate the differences between the outcomes that these models provides. The key point is understanding how well one model performs than the other when trained on same datasets. The evaluation of these models is done based standard evaluation parameters (MAE, MSE, R recall, etc.). The graphs included will also help in determining which model is best for doing time series for stock price prediction.

## II. REVIEW OF RELATED LITERATURE

The time series is the set of quantitative observations arranged in chronological order. Time series analysis has attracted a lot of attention in the past three decades [1]. In the past, it is generally believed that time can be a continuous or discontinuous variable and no comparison exists between the dependent variables [2]. Time series have always been used in the field of econometrics. Jan Tinbergen (1939) devised the first time series econometric model. He also started the scientific research program on the basis of experienced econometrics. At that time, it was hardly considered that chronologically ordered observations might depend on each other. The dominating assumption was that, according to the classical linear regression model, the surplus of the estimated equations is randomly determined and is independent of each other [3].

The fundamental goal of building a time series model is the same as building a precision model that provides a value nearly equal to the values present in the series. From a statistical point of view, time series are the recordings of aleatory processes which vary over time. The most distinguishing feature of time series is that the distribution of observation at a specific point conditional on the previous value of the series depends on the outcome of those previous observations, simply making the outcomes independent. The recordings can be a continuous pattern or a set of discrete values. There are two main types of time series patterns that exist, stationary and non-stationary. Stationary time series values have statistical properties and moments while non-stationary values are simple recordings with changing statistical properties. Both types of series can be used in time series forecasting models [5] [6].

Time series forecasting model building is done based on the type of dataset used to train models. Stationary datasets are easier to train for prediction than non-stationary datasets. In fact, it is a necessity to convert a non-stationary dataset into a stationary dataset [7]. Models easily understood stationary datasets and extract information more efficiently from them. There are certain methods devised to convert non-stationary datasets into stationary datasets [8]. The popular methods are the Hilbert-Huang transform, Fourier transform, Dicky Fuller, etc. Hilbert Hhuang method is specially developed for analyzing non-linear and non-stationary data [9].

Naturally, people's habit of forecasting and making predictions is immemorial. Simply forecasting is the name of predicting outcomes of a plan, but in python, the forecasting is done using datasets available and training them using in-built functions. Using time series forecasting is the process of finding possible values for anything using a known data set. Time series is a popular technique in the current era to solve all types of problems, predicting directly affects the decisions and escort towards clearer imagination. Time series forecasting can be done using both machine learning and deep learning models. The models are different based on their working specifications [10].

There are numerous models in python that serve the purpose of predicting values. Supervised machine learning models and deep neural networks are used for prognostications. Supervised machine learning empowers systems with the ability to learn automatically from the data and get better outcomes with experience without being explicitly programmed [11]. Similarly, deep networks have applications in many areas of life including prediction, detection, creation, etc. Deep neural networks model provide better forecasts as compared to machine learning models. When dealing with real world's problems shallow neural networks need to be sufficiently expressive to predict the task optimally while deep neural networks (DNNs) have been proposed as a way of producing more predictive models [12]. These models are sequences of layers in which each layer uses a linear transformation function. When the layers are combined, it constructs a deep neural model. As many functions are included, that automatically enhances the models' prediction ability.

These models have complex backgrounds and seem difficult to build. With the passage of time, there came a great evolution and now these models are readily present in python libraries. Python has libraries with inbuilt forecasting models that can be used to do predictions. LSTM and ARIMA are the two most influential and long-established models. Models are used considering the characteristics of the encountered problem. A traditional LSTM is a sub-type of RNN model that saves previous sequential data as temporal pieces of information [13], [14], [15].

A general LSTM model comprises the cell; each cell has three parts, i.e. Forget gate, Input Gate and Output Gate. These states are affected by both past states and current input with feedback connections. A standard recurrent cell consists of Sigma and Tanh cells. Each memory cell in LSTM has recurrently self-connected linear unit LSTM. These linear

units are called CEC (Constant Error Council). The architecture of LSTM permits it to bridge huge time lags between relevant input events of almost 1000 steps and more. This method is used in the processing of time-series data, in prediction, as well as in the classification of data.

LSTM models work efficiently and have been widely used in various kinds of tasks. The activation function in the output layer determines in which direction the training will lead here [16], [17].

ARIMA (autoregressive integrated moving average) is also a sub-type of RNN networks and has proven immensely useful for prediction [18], [19], and [20]. They have been studied thoroughly and remodeling has been done on ARIMA processes. These models were popularized by George Box and Gwilym Jenkins [21]. That's why ARIMA processes are sometimes known as Box-Jenkins models. They effectively put together everything in a comprehensive manner and the relevant information required to understand and use univariate ARIMA processes. ARIMA models have gained popularity because they can very accurately do short-term forecasts. Sometimes there is some information that cannot be extracted through regression, so ARIMA is used to capture this additional information. In these models, autocorrelation and partial autocorrelation functions are used, as basic instruments, to identify the stationarity of time series [22]. Both LSTM and ARIMA models have their own specific functionalities and their own advantages and disadvantages.

The previous studies showcase the use of ARIMA and LSTM models in different industries. The use of these models is common in finance and commerce markets but they cannot be seen used in trading markets. The literature gives insights that use of LSTM and ARIMA models in prediction problems is very limited. The research is done to discuss the results of implementation of these two models and evaluating the performance models when they are trained on same dataset. The results helps estimate which model is more useful for stock exchange prediction.

### III. METHODOLOGY

Time series forecasting is used to deal with a surprisingly vast set of problems. There are different techniques available to implement time series forecasting, each differs slightly but affects greatly. The process consists of five major steps. As in this paper, High Price prediction is being done using two different models (i.e. LSTM and ARIMA). This five-step methodology is used to implement these two models. Although the general working of all five models is identical, it varies greatly according to the technicalities used in each model.

Fig. 1 shows the general methodology steps which are explained, respectively.



Fig. 1. General methodology flowchart.

A. Understand the Problem

As mentioned earlier, time series data is used to deal with lots of problems. Time series forecasting is the type where certain solutions are speculated by training previous data. The solutions are; predicting stock market reach at a certain time, estimating survey results based on previous data, weather prediction at a certain date, etc. it is important to build a model keeping in view the nature of the problem you are dealing with. Explore the models that can be used according to the given problem. In addition to understanding models, thoroughly understand the techniques used to implement these models in order to choose the best according to the problem.

The purpose of this research is to build two different time series forecasting models in order to proclaim which is the best model for predicting high prices. To do so, there are two best-performing models (LSTM & ARIMA) that have been working exceptionally well while dealing with prediction problems.

B. Data Gathering / Pre-processing

After understanding the nature of the problem, the next step is to determine the type of data needed in order to train the models. The data should be the cross-sectional type of data. This type has one dependable variable that confides in numerous independent variables. In such data types, the small-scale or aggregate entity is observed at different points in time. In a nutshell, this is the data of different entities collected at the same time.

In this paper, data from Mulkia Gulf Real Estate from Saudi Exchange datasets are used. The data is extensive to cause models to overfit, so certain sheets (i.e. Fig. 2; sheet 1 & Fig. 3; sheet 7) are chosen to train models. It is a cross-section of data perfectly meeting the criteria of the needed data. The purpose of using this data is because the data possess similarities that are consonant with the prediction problem. After training models on this data the trained model can be used to predict the high price value. This data accurately meets the criteria of needed data to predict the highest price value.

After data has been selected, the next step is to pre-process this data. As mentioned earlier, the data type used is non-stationary and cross-sectional data. The models used can train on stationary data more efficiently, so it is necessary to convert this data into stationary data. There are certain methods available to convert non-stationary data into stationary. ADF (Argumented Dicky Fuller) method is used to transform the data into non-stationary. In python, ADF can be imported from the Stats model library as statsmodels.tsa.stattools.adfullers.

C. Exploratory Analysis

Exploratory analysis is the process of analyzing the relationships between variables that exist in raw data. These initial relationships help to understand the nature of the data and how accurately the desired information will be extracted from it. It's mandatory to do exploratory analysis independent of the type of problem, it's mandatory to explore the data first. It's also called preliminary analysis of data, where you plot the data in its original form to find certain structures. When

preliminary analysis is done, check the validity of measures, and point out any outliers. It also helps in examining the weightage of different variables to evaluate the effectiveness of certain manipulative variables, etc. The major thing that should be implemented while doing preliminary analysis is cleaning disrupted data, checking the data for null values, etc.

Fig. 4 shows that the data has been scaled properly and is ready to train for the prediction models. It also shows the regression patterns of various variables alongside dates.

D. Choosing Libraries and Training Models

In Python, there are thousands of libraries and in-build models available that deal with machine learning and deep learning problems. Python libraries are a set of specific functions put together in a single file. The purpose to make these libraries is to assist coders in doing obvious steps. There is a whole set of libraries available for different domains of artificial intelligence. The problem explained in this research paper is time series forecasting, which is a type of machine learning and deep learning problem. The libraries which come in handy while dealing with such problems are sklearn, tensorflow, numpy, etc.

Company	ID	Date	Cross Price	Open Price	Highst Price	Lowest price	VOLUME	%CHANGE	
Mulkia REIT	4330	24-May-22	10.74	10.72	10.78	10.69	108.62K	-0.18%	
Mulkia REIT	4330	23-May-22	10.76	10.74	10.76	10.75	16.75Z	-0.37%	
Mulkia REIT	4330	20-May-22	10.8	10.8	10.8	10.82	18.1415.74K	-0.16%	
Mulkia REIT	4330	19-May-22	10.74	10.82	10.82	10.82	10.69	210.56K	-1.47%
Mulkia REIT	4330	18-May-22	10.9	10.86	10.9	10.7	10.7	265.60K	1.11%
Mulkia REIT	4330	17-May-22	10.78	10.78	10.88	10.88	10.82	310.69K	0.95%
Mulkia REIT	4330	15-May-22	10.84	10.74	10.86	10.74	10.74	158.90K	-0.55%
Mulkia REIT	4330	14-May-22	10.74	10.88	10.88	10.74	10.74	248.73K	0.19%
Mulkia REIT	4330	11-May-22	10.72	10.8	10.88	10.88	10.74	430.39K	-0.74%
Mulkia REIT	4330	10-May-22	10.8	10.86	10.94	10.84	10.84	107.65K	-1.46%
Mulkia REIT	4330	9-May-22	10.96	10.92	10.96	10.96	10.96	539.14K	0.18%
Mulkia REIT	4330	8-May-22	10.94	10.98	11	11	10.94	107.30K	-0.16%
Mulkia REIT	4330	6-May-22	10.96	11	11.04	11	10.94	206.30K	0.65%
Mulkia REIT	4330	27-Apr-22	10.96	10.96	11	11	10.96	120K	-0.16%
Mulkia REIT	4330	26-Apr-22	10.96	11.06	11.06	11.06	10.94	150.99K	0.16%
Mulkia REIT	4330	25-Apr-22	11	11	11	11	10.94	217.40K	0.36%
Mulkia REIT	4330	24-Apr-22	10.96	10.9	10.96	10.94	10.94	207.10K	0.92%
Mulkia REIT	4330	21-Apr-22	10.96	11	11	11	10.94	411.93K	1.27%
Mulkia REIT	4330	20-Apr-22	11	11.1	11.1	11.1	11.1	115.69K	-0.60%
Mulkia REIT	4330	19-Apr-22	11	11.1	11.2	11.1	10.88	131K	1.68%
Mulkia REIT	4330	18-Apr-22	11.12	11.08	11.16	11.08	11.08	216.69K	0.36%
Mulkia REIT	4330	17-Apr-22	11.68	11.14	11.16	11.16	10.88	236.30K	-0.54%
Mulkia REIT	4330	16-Apr-22	11.14	10.9	11.16	11.16	10.9	513.96K	2.26%
Mulkia REIT	4330	13-Apr-22	10.9	10.76	10.92	10.92	10.74	309.79K	1.49%
Mulkia REIT	4330	12-Apr-22	10.74	10.76	10.78	10.78	10.82	327.68K	0.37%
Mulkia REIT	4330	11-Apr-22	10.7	10.76	10.78	10.78	10.7	238.86K	-0.19%
Mulkia REIT	4330	10-Apr-22	10.72	10.76	10.78	10.78	10.72	282.30K	-0.15%

Fig. 2. Mulkia gulf real estate from Saudi exchange sheet 1.

Company	ID	Date	Cross Price	Open Price	Highst Price	Lowest price	VOLUME	% CHANGE
Mulkia Gulf Real Estate REIT	4336	24-May-22	9.47	9.48	9.5	9.44	71.56K	-0.21%
Mulkia Gulf Real Estate REIT	4336	23-May-22	9.49	9.5	9.5	9.46	84.57K	-0.11%
Mulkia Gulf Real Estate REIT	4336	22-May-22	9.5	9.52	9.52	9.47	101.12K	-0.21%
Mulkia Gulf Real Estate REIT	4336	19-May-22	9.52	9.54	9.54	9.46	244.65K	-0.21%
Mulkia Gulf Real Estate REIT	4336	18-May-22	9.54	9.53	9.54	9.51	115.39K	-0.10%
Mulkia Gulf Real Estate REIT	4336	17-May-22	9.55	9.6	9.6	9.54	160.85K	-0.31%
Mulkia Gulf Real Estate REIT	4336	16-May-22	9.58	9.59	9.59	9.56	73.21K	-0.10%
Mulkia Gulf Real Estate REIT	4336	15-May-22	9.59	9.59	9.6	9.58	67.39K	0.10%
Mulkia Gulf Real Estate REIT	4336	12-May-22	9.58	9.6	9.6	9.58	182.25K	-0.21%
Mulkia Gulf Real Estate REIT	4336	11-May-22	9.6	9.7	9.71	9.55	613.00K	-1.64%
Mulkia Gulf Real Estate REIT	4336	10-May-22	9.76	9.82	9.82	9.76	272.83K	-0.37%
Mulkia Gulf Real Estate REIT	4336	9-May-22	9.79	9.78	9.8	9.78	192.86K	0.10%
Mulkia Gulf Real Estate REIT	4336	8-May-22	9.78	9.81	9.82	9.78	198.64K	-0.20%
Mulkia Gulf Real Estate REIT	4336	28-Apr-22	9.8	9.86	9.86	9.79	255.77K	-0.41%
Mulkia Gulf Real Estate REIT	4336	27-Apr-22	9.84	9.83	9.84	9.8	140.59K	0.20%
Mulkia Gulf Real Estate REIT	4336	26-Apr-22	9.82	9.82	9.84	9.8	99.93K	0.00%
Mulkia Gulf Real Estate REIT	4336	25-Apr-22	9.82	9.87	9.87	9.8	134.91K	-0.51%
Mulkia Gulf Real Estate REIT	4336	24-Apr-22	9.87	9.94	9.94	9.86	168.38K	-0.10%
Mulkia Gulf Real Estate REIT	4336	21-Apr-22	9.88	9.91	9.91	9.88	112.54K	-0.10%
Mulkia Gulf Real Estate REIT	4336	20-Apr-22	9.89	9.92	9.92	9.87	126.73K	-0.30%
Mulkia Gulf Real Estate REIT	4336	19-Apr-22	9.92	9.92	9.92	9.9	101.77K	-0.10%
Mulkia Gulf Real Estate REIT	4336	18-Apr-22	9.93	9.98	9.98	9.85	165.47K	-0.40%
Mulkia Gulf Real Estate REIT	4336	17-Apr-22	9.97	9.99	10	9.97	355.29K	-0.10%
Mulkia Gulf Real Estate REIT	4336	14-Apr-22	9.98	9.99	9.99	9.97	71.16K	-0.10%
Mulkia Gulf Real Estate REIT	4336	13-Apr-22	9.99	9.97	9.99	9.91	224.73K	0.10%
Mulkia Gulf Real Estate REIT	4336	12-Apr-22	9.98	9.95	9.98	9.91	212.86K	0.60%
Mulkia Gulf Real Estate REIT	4336	11-Apr-22	9.92	9.92	9.92	9.92	20.18K	0.00%
Mulkia Gulf Real Estate REIT	4336	10-Apr-22	9.92	9.98	9.98	9.89	95.15K	0.00%

Fig. 3. Mulkia gulf real estate from Saudi exchange sheet 7.

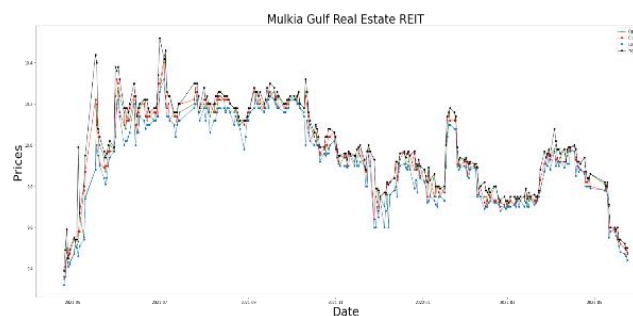


Fig. 4. Exploratory analysis.



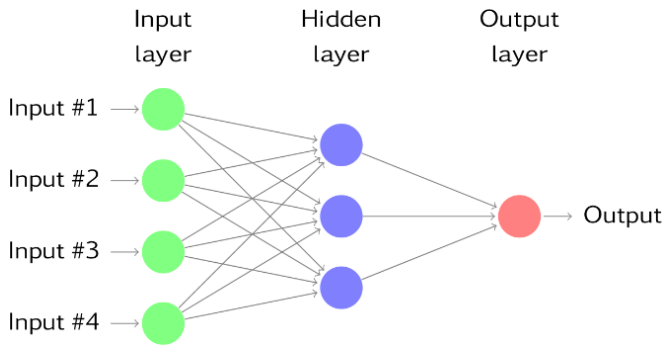


Fig. 5. Basic deep learning neural network.

These libraries are designed according to a specific set of rules that fit their receptive type of problem. For preprocessing of this data; numpy, sklearn, tensorflow, and keras are used. Similarly, there are some in-built forecasting models that are available in python. Use these models suitably for respective problems. The models serve specific purposes; they have been created according to some mathematical ordinances. According to the problem type choose these rules and embed them into deep learning models. A typical deep learning model consists of an input layer, multiple hidden layers (where most of the work is done), and the output layer. Fig. 5 shows the general representation of a deep learning neural network model.

Now keeping in view the nature of the problem, use LSTM and Auto-Regressive models.

### E. Evaluating Models

Evaluation is done by measuring evaluation parameters. The three basic parameters that are used in measuring a neural model's reliability in time series forecasting problems are Accuracy, Precision, and Recall. In addition to these parameters, some auxiliary parameters are available to evaluate regression results. The added parameters are MSE (Mean Square Error), MAE (Mean Absolute error), MAPE (Mean Absolute percentage), and MDAPE (Median Absolute Error Average and Percentage). These metrics collectively help provide an explanation regarding the mistakes made unknowingly.

## IV. IMPLEMENTING TIME SERIES MODELS

In this research paper, the two best time series forecasting models are used that are already available in python. These models will be trained on two datasets from Mulkia Gulf Real Estate from Saudi Exchange. Sheet 1 & Sheet 7 of these datasets will be used to train the models. The aim is to determine which model is the best model for respective targeted industry. The problem is predicting the highest profits values. The trained models will be attested, and finding will be evaluates based on evaluation parameters to get exquisite results. The general methodology explained above will be followed. The models are embedded one by one into the neural networks. After embedding the models, the pre-processed datasets will be provided and will run in the code for the training of models. The general method proposed to train the models is nearly identical. They only differentiate based on the functionality attributes, and the technical

differences, i.e. number of hidden layers used, the evaluation parameters used, and the dataset preparations, the working background. Detailed information on these python models is provided below.

### A. Long Short-Term Memory Loss (LSTM)

As the prediction model should remember long-lasting events from part, LSTM is the first choice to use in this paper. LSTM is an inbuilt python function that can be imported from TensorFlow. LSTM is the branch of recurring neural networks. It was seen that there weren't any RNN structures that can do backpropagation of long intervals, so to solve such difficulties LSTM was proposed. LSTM has two gate unit cells that open and close to the information flow within each memory cell, i.e. Packs of information between time lags [23]. It requires the data used to be in a definite shape. The dimension should be equivalent, and the data should be properly cleaned, integrated, and scaled. The commonly used activation functions for LSTM-based regression problems are Sigmoid and Tanh. Tanh has proven to be very effective in dealing with vanishing gradients.

The functioning of the LSTM model is explained in Fig. 6.

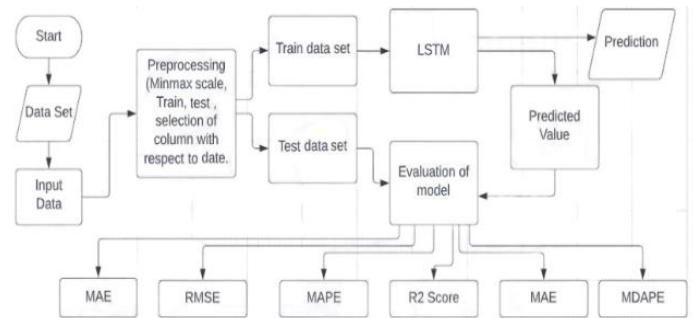


Fig. 6. Flowchart for LSTM working.

As said earlier, LSTM has the ability to remember information for the long term. This is done by remembering the previous output and combining it with the current one. To understand the working of LSTMs better, there is a need to fathom the mathematics behind them.

Divide the working of LSTM neural networks into four stages. In the first stage, the model decides whether to forget or remember information in the previous cell. This value is decided by doing the following calculation:

$$\text{Forgetting value } f_t = \sigma [ W_f (h_{t-1}, X_t) + b_f ]$$

Where  $\sigma$  is the activation used in the input layer,  $W_f$  and  $B_f$  are the weight and bias vectors,  $h_{t-1}$  is the output at time  $t-1$ , and  $X_t$  is the input vector at time  $t$ . If the forgetting value is equal to zero then this means that the previous value has been forgotten.

In the second stage, the model decides which value will get stored in the next cell state. To do so, sigmoid and tanh layers are used. The sigmoid layer chooses the pieces of information which need to be updated, and the layers decide on a second option value. By combing these two values, the models create new values to update the cell state. The formula to calculate sigmoid and tanh values are:

Input gate value  $i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i)$

Updated value  $C_t = \tanh(WC \cdot [h_{t-1}, X_t] + b_C)$

$\sigma$  represents the sigmoid-shaped function, it is the input gate value and  $C$  is the updated value at time  $t$ .

In the third stage, the new cell state  $NC_t$  is obtained using  $f_t$  and its value. The equation below shows how to calculate this  $NC_t$  value,

$$C_t = f_t [C_{t-1} + i_t] e$$

In the final stage of LSTM, it is determined which value will get considers as output. Use the sigmoid equation to determine which cell state will be the output; the value is processed through  $\tanh$  to get a value between 1 and -1.

The inbuilt model present in python work is according to the above-explained structure. The models can be improved further by choosing the correct values of weights and bias values, activation functions, and the number of hidden layers.

### B. Autoregressive Integrated Moving Average (ARIMA)

ARIMA forecasts temporal dependencies using only historical values. The data used for Autoregressive models are prepared differently from LSTM. In addition to necessary preprocessing steps, the AR model's data needs to be stationary. Simply put, data is stationary when its numeric properties do not change over time. From a mathematical perspective, it refers to the data whose Mean and Variance will not depend on time. If the data doesn't meet the properties of the stationary dataset, you can do a series transformation to make it stationary.

ARIMA model is a combination of two models Autoregressive (AR) and Moving Average (MA), integration (I) is applied at least once to make the data set stationary.

In the AR part of the model future values are predicted using the lags from the data values. The general equation AR model is:

$$AR(p): x_t = \alpha + \sum_{i=1}^p \beta_i x_{t-i} + \epsilon$$

The Moving Average is the part of the model where value is forecasted using the forecasting error differences is calculated while making predictions. The general form of MA equations is:

$$x_t = \mu + \sum_{i=1}^q \Phi_i \epsilon_{t-1}$$

Prediction is done by combining all three orders and getting an estimation of how to quickly fit the model. Some standard denotations are used to represent the above three, i.e.  $p$ ,  $d$ , and  $q$ ; " $p$ " is the number of observations included in the model, " $d$ " is the number of times differentiating the raw observations, " $q$ " is the number of moving average size. To find these parameters, first fetch a Correlation and Partial Correlational graph from the dataset.

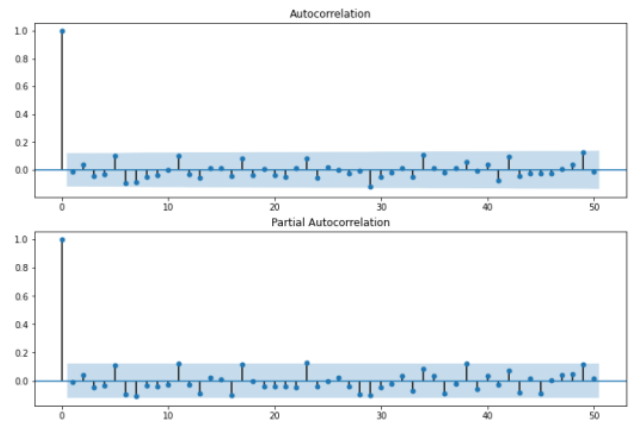


Fig. 7. Autocorrelation graph of sheet 1 data.

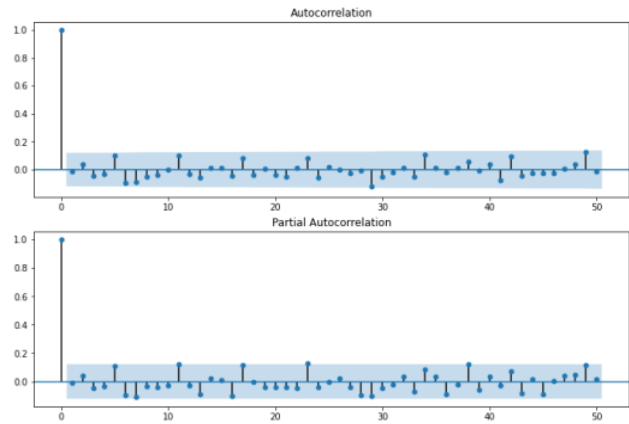


Fig. 8. Autocorrelation graph of sheet 7 data.

Fig. 7 and 8 can be used to roughly estimate the value of  $p$ ,  $d$ , and  $q$ . The integer value of  $p$  can be obtained from the cut-off edge points of PAC graphs. Similarly, the value of  $q$  can be obtained using an AC graph. If the graph does not represent steady cut-off points, use in-built PACF and CAF functions.

Fig. 9 shows how the ARIMA works and on which parameters its working can be evaluated.

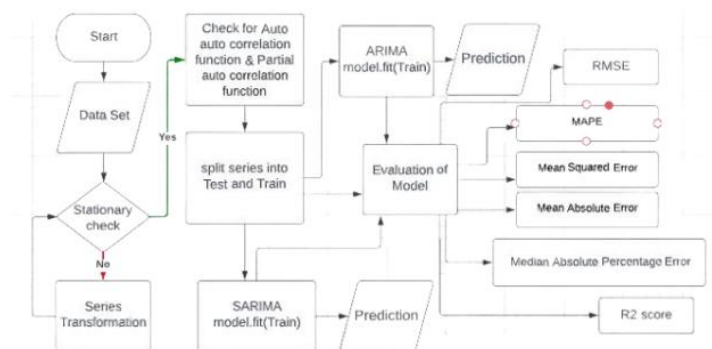


Fig. 9. Flowchart for ARIMA working.

## V. EVALUATION PARAMETERS FOR MODEL'S PERFORMANCE

Taking note of the significance of time series models, it is very important to examine the results of these models. For specific industries, the targeted values of forecasting models are the milestones for making decisions [9]. Artificial intelligence learning models are always trained using real-world data. They are prepared to use for actual problems, so they are usually tested using the same type of data.

Keeping in view the importance of time series models, scientists have devised a number of ways to test the audacity of the models used. In this part, the performance parameters the will be used to evaluate the models will be discussed, and how best model among them will be picked. Following are the parameters used for testing model's prediction validation.

### A. R Square Score - R2 Score

It's a statistical value that is used in a prediction model to find the extent of difference between dependent variables that can be made sense of by the independent factor. In simple words, R2 is how well the testing data fits the trained regression model. The general formula used to calculate the R2 score of the training models is.

$R^2 = 1 - \frac{\text{sum of squares due to regressions}}{\text{total sum of squares}}$

The R2 score can only be calculated for both. The R2 values for both models are displayed in Table I.

TABLE I. R2 VALUES

R2 Score	LSTM	ARIMA
Sheet 1	-0.2142	0.826684
Sheet 7	-0.3214	0.7389812

### B. Mean Square Error - MSE

It is the mean square of differences between the trained model values and tested values. It squares the values of the differences in order to remove the negative sign and increase the weight of the larger values. The formula used to calculate the MSE is

$$MSE = \sum (X_t - X'_t)^2 / n$$

n = Number of items forecasted

t = Time period values

X = actual values of the dependable variable (In the case, actual values of the highest prices)

X' = Prediction values of the dependable Variable

In order to estimate the performance of the model, calculate MSE, the smaller a value is the better the model prediction. The MSE value for the ARIMA model (Sheet 1 & Sheet 2; 0.033151 and 0.015704).

### C. Root Mean Square Error - RMSE

RMSE is obtained by taking the square root of MSE equations. Root square is added in order to return the MSE value to become consistent, not compromising the ability to

penalize error. There are certain models that contain inconsistent but larger errors. To calculate the possibility of occurring such an error, compare RMSE with MAE to see if such errors are present. The RMSE value for LSTM is (Sheet 1 & Sheet 2; 0.088/0.474 and 0.130/0.232).

### D. Mean Absolute Error - MAE

It is the mean of the absolute difference between the observed value, and the real observation value. In other words, it tells how much larger the difference between actual and predicted values which can be expected from the model. The smaller the value of MAE, the better the model will work. If the value is zero, the model can predict future values accurately. Models are compared on the basis of MAE values such as; a model with a smaller MAE will be considered best among all. MAE simply calculates the error; it can't identify the weightage of individual values. The general equation used to measure MAE is:

$$MAE = \sum (X - X') / n$$

The MAE value for ARIMA is (Sheet 1 & Sheet 2; 0.111602 and 0.071160).

### E. Mean Absolute Percentage Error - MAPE

MAPE is obtained by measuring the percentage of MAE to the real values, (In this case it's X - Real values of Prices. Mainly find MAPE, when data is without Zeroes and Extreme values. Read this error the same as MAE which means the lower MAPE value model is the best model. The MAPE value for ARIMA is (Sheet 1 & Sheet 2; 0.011073 and 0.007107).

### F. Median Absolute Percentage Error - MDAPE

A little modification is done in MAPE - Mean Absolute percentage error in order to get MdAPE. The median of the MAE is found by arranging the values from the smallest to the largest; then the middle value only if it's even is picked as MdAPE.

MdAPE is recommended for evaluation when you are dealing with ARIMA models. MDAPE evaluation is done for ARIMA models. MdAE Percentage is interpreted as good when it's between 10% and 20%. The RMSE value for ARIMA is (Sheet 1 & Sheet 2; 0.630533 and 0.333124).

In addition to all the parameters mentioned above, also study the avg\_loss and val\_loss graphs to evaluate models like LSTM. These additional parameters help to evaluate and compare the two models better.

## VI. RESULTS

The purpose of the research is to know which model is the best-performing model for the Highest Price value prediction among the two that have trained using different Python libraries. The results will comprehend using graphs combined with the evaluation parameters explained above.

As mentioned, two sets of Data have been used; Sheet 1 & Sheet 7 from the dataset of Mulkia Gulf Real Estate available at Saudi Exchanges. The first model trained on this data is Long Short Term Memory (LSTM). The evaluation graph shows the regression between actual trained and prediction train values from sheet 1 to sheet 7 as follows:



Fig. 10. Sheet 1 LSTM prediction.

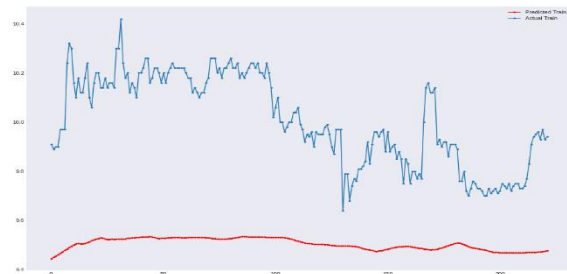


Fig. 11. Sheet 7 LSTM prediction.

Fig. 10 and 11 shows the regression differences present between actual value and predicted values. The LSTM model is evaluated using RMSE, avg\_loss, and val\_loss. The RMSE value for sheet 1 is 0.088 & for sheet 7 is 0.130. The average and validation losses are between 0.08 and 0.17.

The next model trained was the Autoregressive Integrated Moving Average (ARIMA) model made by importing ARIMA from statsmodels.tsa library from python. The graphs indicating the model results for real and predicted values are:

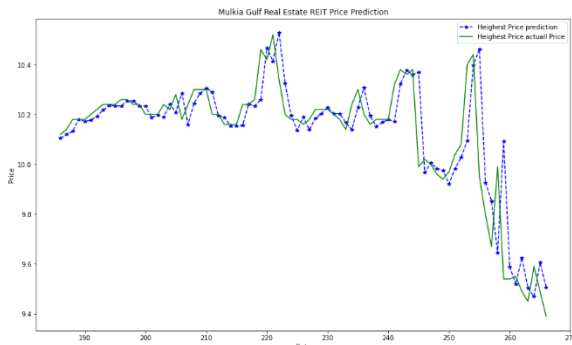


Fig. 12. Sheet 1 ARIMA prediction.

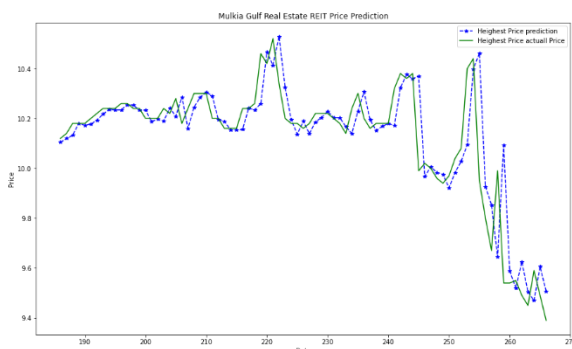


Fig. 13. Sheet 1 ARIMA prediction.

The visualization shows (Fig. 12 & 13) shows very little difference between the actual value and the predicted values from ARIMA.

The mathematical evaluation of ARIMA was done using MAE, MAPE, MdAPE, and MSE. You can read those values from tables. The overall values gave an idea that ARIMA can predict the future highest price more accurately than LSTM.

## VII. CONCLUSIONS

In this paper, the time series forecasting problem is explained. In this paper, the highest price prediction using two python in-built models (i.e. LSTM & ARIMA) is done. The data is trained on the Mulikia Gulf Real State dataset. The dataset is pre-processed and trained on deep-learning models. After training evaluation of models is done using evaluation parameters; MAE, MSE, RMSE, val\_loss, accuracy, R2 score, etc. The values of these parameters collectively determine which model performed the best. These parameters indicate that the ARIMA model can predict the highest price more accurately. When viewed from the graphical point of view, there is very little difference between actual values and predicted values in Fig. 12 and 13, this shows that ARIMA can predict more precisely as compared to LSTMs. Overall, it was concluded that for stock price prediction, ARIMA models can perform better than LSTM Models.

## REFERENCES

- [1] Kam, Kin Ming. *Stationary and non-stationary time series prediction using state space model and pattern-based approach*. The University of Texas at Arlington, 2014.
- [2] Hyndman, Rob J., and George Athanasopoulos. *Forecasting: principles and practice*. texts, 2018.
- [3] Kirchgässner, Gebhard, Jürgen Wolters, and Uwe Hassler. *Introduction to modern time series analysis*. Springer Science & Business Media, 2012.
- [4] Charlton, M. & Caimo, A. *Time Series Analysis*. European Spatial Planning Observation Network (ESPON). 2012.
- [5] Wei, William WS. "Time series analysis." (2013).
- [6] Parzen, Emanuel. "An approach to time series analysis." *The Annals of Mathematical Statistics* 32.4 (1961): 951-989.
- [7] Manuca, Radu, and Robert Savit. "Stationarity and nonstationarity in time series analysis." *Physica D: Nonlinear Phenomena* 99.2-3 (1996): 134-161.
- [8] Manual, Solution. "Data Mining: Concepts and Techniques".
- [9] Huang, Norden E., et al. "Applications of Hilbert–Huang transform to non-stationary financial time series analysis." *Applied stochastic models in business and industry* 19.3 (2003): 245-268.
- [10] Yaffee, Robert A., and Monnie McGee. *An introduction to time series analysis and forecasting: with applications of SAS® and SPSS®*. Elsevier, 2000.
- [11] Choudhary, Rishabh, and Hemant Kumar Gianey. "Comprehensive review on supervised machine learning algorithms." *2017 International Conference on Machine Learning and Data Science (MLDS)*. IEEE, 2017.
- [12] Samek, Wojciech, et al. "Explaining deep neural networks and beyond: A review of methods and applications." *Proceedings of the IEEE* 109.3 (2021): 247-278.
- [13] Yu, Yong, et al. "A review of recurrent neural networks: LSTM cells and network architectures." *Neural computation* 31.7 (2019): 1235-1270.
- [14] Gers, Felix A., Nicol N. Schraudolph, and Jürgen Schmidhuber. "Learning precise timing with LSTM recurrent networks." *Journal of machine learning research* 3.Aug (2002): 115-143.

- [15] Pascanu, Razvan, Tomas Mikolov, and Yoshua Bengio. "On the difficulty of training recurrent neural networks." *International conference on machine learning*. PMLR, 2013.
- [16] Elsworth, Steven, and Stefan Güttel. "Time series forecasting using LSTM networks: A symbolic approach." *arXiv preprint arXiv:2003.05672* (2020).
- [17] Yu, Yong, et al. "A review of recurrent neural networks: LSTM cells and network architectures." *Neural computation* 31.7 (2019): 1235-1270.
- [18] Jakaša, Tina, Ivan Andročec, and Petar Sprčić. "Electricity price forecasting—ARIMA model approach." *2011 8th International Conference on the European Energy Market (EEM)*. IEEE, 2011.
- [19] Contreras, Javier, et al. "ARIMA models to predict next-day electricity prices." *IEEE transactions on power systems* 18.3 (2003): 1014-1020.
- [20] Meyler, Aidan, Geoff Kenny, and Terry Quinn. "Forecasting Irish inflation using ARIMA models." (1998): 1-48.
- [21] Box, GEORGE EP, Gwilym M. Jenkins, and G. Reinsel. "Time series analysis: forecasting and control Holden-day San Francisco." *BoxTime Series Analysis: Forecasting and Control Holden Day1970* (1970).
- [22] Contreras, Javier, et al. "ARIMA models to predict next-day electricity prices." *IEEE transactions on power systems* 18.3 (2003): 1014-1020.
- [23] Sepp Hochreiter, Jürgen Schmidhuber; Long Short-Term Memory. *Neural Comput* 1997; 9 (8):1735–1780.doi: <https://doi.org/10.1162/neco.1997.9.8.1735>.

# Enhancing the Intrusion Detection Efficiency using a Partitioning-based Recursive Feature Elimination in Big Cloud Environment

Hesham M. Elmasry<sup>1</sup>, Ayman E. Khedr<sup>2</sup>, Hatem M. Abdelkader<sup>3</sup>

Management Information Systems Department-Faculty of Commerce and Business Administration, Future University in Egypt (FUE), Cairo, Egypt<sup>1</sup>

Information Systems Department-Faculty of Computers and Information Technology, Future University in Egypt (FUE) Cairo, Egypt<sup>2</sup>

Information Systems Department-Faculty of Computers and Information, Menoufia University, Menoufia, Egypt<sup>3</sup>

**Abstract**—In the era of cloud computing, the effectiveness of utilizing supervised machine-learning-based intrusion detection models for categorizing and detecting malicious network attacks depends on the preparation, extraction, and selection of the optimal subset of features from the dataset. Therefore, before beginning the training phase of the machine learning classifier models, it is required to remove redundant data, manage missing values, extract statistical features from the dataset, and choose the most valuable and appropriate attributes using the Python Jupyter Notebook. In this study, partitioning-based recursive feature elimination (PRFE) method was suggested to decrease the complexity space and training time for machine learning models while increasing the accuracy rate of detecting malicious attacks. On the information security and object technology cloud intrusion dataset (ISOT-CID), some of the most popular supervised machine learning classification techniques, including support vector machines (SVM) and decision trees (DT), have been assessed using the suggested PRFE technique. In comparison to some of the most popular filter and wrapper-based feature selection strategies, the results of the practical experiments demonstrated an improvement in accuracy, recall, F-score, and precision rate after using the PRFE technique on the ISOT-CID dataset. Additionally, the time required to train the machine-learning models was reduced.

**Keywords**—Machine learning models; big cloud environment; intrusion detection system (IDS); Jupyter Notebook; feature selection; ISOT-CID introduction

## I. INTRODUCTION

In the era of cloud computing and with the steady growth in the volume of transmitted and received data, machine learning models have emerged as one of the most significant contemporary techniques used to recognize and categorize dangerous assaults from network traffic. Preprocessing techniques on the data are therefore necessary in order to increase the precision and effectiveness of these models. A fundamental set of sub processes known as data preparation comprises steps including deleting duplicate data, filling in missing values, and turning some categorical data into numerical data so that machine learning models can interpret it [1], [2]. In a machine-learning process, incoming data is analyzed by computers to create patterns that foretell learning outcomes with a minimum of human input [3]. Based on how

the learning algorithm is implemented, machine learning models can be divided into three groups.

The supervised machine learning (SML) model is used when the data available for the training phase is labelled, which means that some dataset attributes contain the correct answer that will be used at the end of the learning process to evaluate the final outputs. This model can be developed using either classification or regression algorithms [4], [5]. When dealing with a dataset that lacks labelled features, the unsupervised machine learning (UML) model is used and relied on; the model instead relies on trial and error to evaluate the learning process's outcomes. Moreover, this model can be developed using clustering algorithms [6]. While the reinforcement machine learning (RML) model evaluates the outcomes of the learning process based on the existence of an entity that performs a set of actions in a specific environment, a reward is given if the action matches the desired result. This model can be done with value-based, policy-based, and model-based algorithms [7].

One of the key elements influencing how well supervised machine learning models can detect and categorize harmful intrusions is feature engineering [8]. This can be done by selecting the dataset's most significant and connected features to the model outputs, a process known as feature selection, and then creating a new feature from the already accessible ISOT-CID dataset, a process known as feature extraction [9], [10].

In the feature selection phase, duplicated features and features that were not related to the outcomes of the learning process were excluded from the ISOT-CID dataset. The focus is only on the features that are most influential in building the detection model and are related to the results of the learning process, which would reduce the time required in the data training process and improve the quality of the outputs. The methods for selecting features can be divided into three types. In the filtering method, the degree of variance is calculated for each feature in the dataset, and higher or equal features are selected by the user based on a predetermined variance threshold [11]. One disadvantage of this method is that it does not consider the relationship between the selected features and target variables.

The wrapper technique uses a sophisticated search algorithm to analyze every feature combination in the dataset, then uses a machine-learning algorithm to evaluate the learning outcomes and choose the feature set that produces the best output. The high rate of classification accuracy for malicious attacks is one of this method's key benefits in terms of selecting the best features. The exorbitant expense and complexity of this technology are also disadvantages. Forward selection (FS), backward selection (BS), and recursive feature elimination (RFE) are three of the most significant algorithms utilized in this strategy [12]. The filter and wrapper methods' issues are addressed by the hybrid approach. There are two sections to the process. The features of the dataset were first created using a filtering technique. In the subsequent step, wrapper techniques were used to select the best features. Two of the most crucial algorithms used in this method are random forest importance (RFI) and LASSO regularization (LR) [13], [14].

The remainder of this paper is organized as follows: The literature review is summarized in Section II. Section III presents the research methodology and the five steps involved in this investigation. A detailed description of the ISOT-CID datasets is provided, along with information on the proposed partitioning-based recursive feature elimination (PRFE) technique, model flowchart, algorithm, performance metrics, research findings, and building ML classifier models. Section IV describes the discussion of results. The final Section V of this paper presents conclusions and suggestions for future research.

## II. LITERATURE REVIEW

This section consists of two main parts. The first part focuses on reviewing and analyzing some of the previous work on supervised machine learning classifier algorithms, such as support vector machines, decision trees, naive Bayes, and k-nearest neighbor algorithms, and their improvements. In addition, the main limitations of each approach were identified. In the second part, we look back at some of the previous studies on feature selection techniques and analyze them. We explain the main improvements and limitations of each technique.

### A. Machine Learning Classifier Algorithms

Without human interaction, intrusion detection systems can recognize new assaults using machine learning (ML). The IDS is able to modify its execution plan by using ML and taking into account recently acquired data. The two main categories of learning strategies are supervised and unsupervised strategies. In supervised learning, examples with input and output labels provided during training are used to "train" algorithms [15]. The unsupervised learning algorithms are allowed to make their own interpretations of the data because the training dataset is empty of labelled data. Unsupervised learning employs clustering and association algorithms to find patterns and distinctions in the data [16].

Peng et al. suggested using supervised machine-learning methods to categorize harmful attacks in a cloud environment and develop a decision tree-based model for intrusion detection. In order to guarantee the efficacy, excellence, and

accuracy of the proposed models for categorizing hostile assaults, researchers have relied on a variety of preprocessing methods to prepare and clean enormous datasets. The suggested model was found to be more efficient and effective than naive Bayesian and k-nearest neighbor models in laboratory trials using the Knowledge Discovery and Data Mining (KDDCUP99) dataset. The decision tree's training duration is not ideal, though, and this is its biggest drawback. Additionally, only the k-nearest neighbor and naive Bayesian models were contrasted with the decision tree model [17].

Using the Apache Spark machine learning library, Belouch et al. conducted a comparison study to assess the effectiveness and detection accuracy of support vector machines, random forests, decision, and naive Bayes algorithms (MLLIB). The results of the lab tests performed on the UNSW-NB15 dataset demonstrated the random forest algorithm's efficiency and effectiveness in comparison to other models. The greatest issue is that the model takes a long time to create and train because the feature selection technique isn't used [18].

Belavagi and Muniyal offered classification and predictive models for intrusion detection using machine learning classification techniques as logistic regression, support vector machine, naive Bayes, and random forest (RF). The techniques are evaluated using the Network Security Laboratory- Knowledge Discovery in Databases (NSL-KDD) dataset. The testing findings demonstrated that, with a peak value of 99%, the Random Forest classifier outperformed the other techniques in all criteria. However, the use of feature selection strategies to choose the best features from the dataset in order to minimize dimensionality is not examined in this paper [19].

An intrusion detection-based big data model was suggested by Azeroual and Nikiforova using unsupervised machine learning and the K-means clustering technique. The correlation-based filter method is used by the author to select the attributes that have the greatest impact on the results of the learning process. The Synchro Phasor dataset used in the laboratory tests revealed a high degree of classification accuracy for harmful attacks. However, the fundamental issue is that the lack of test support in the Apache Spark framework prevented the authors from comparing the suggested model to other solutions [20].

Souhail et al. suggested a two-stage network-based IDS (NIDS) technique to recognize network threats. The proposed approach combines LR, gradient boost machine (GBM), support vector machine (SVM), recursive feature elimination (RFE), and random forest feature selection methods for the complete UNSW-NB15 dataset. The results showed that the accuracy rate of multi-classifiers using decision trees was about 86.04%. Due to the usage of the recursive feature elimination-based feature selection technique, the key restriction is the amount of time needed to create and train the model [21].

A detection framework using an ML model was proposed by Alshammari and Aldrabi to feed IDS and detect abnormal network traffic in cloud environments. An ISOT-CID dataset containing both malicious and normal traffic is used in this detection method. Six machine-learning models were trained

using this dataset, and they were then tested using split- and cross-validation techniques. Only two of the results were satisfactory, but the other four were accurate enough to be useful. The model's reliance on a large dataset or considerable dataset, which has an impact on how well the system is fitted and evaluated, is the biggest drawback, though [22].

### B. Feature Selection Algorithms

Using evolutionary algorithms and support vector machines, Ashahri et al. suggested an embedding-based feature selection technique to minimize the number of dataset features from 45 to 10. The ten traits that had been chosen were then divided into three groups based on their level of significance in the following stage. Laboratory tests revealed that the suggested hybrid algorithm has a true-positive value of 0.973 and a false-positive value of 0.017 [23].

Based on 70% of the DDOS dataset from NSL-KDD, Mohammed and Gyasi suggested an intrusion detection system for distributed denial-of-service (DDOS). Random forest (RF) and multilayer perceptron (MLP) were utilized for the detection tasks, and recursive feature elimination (RFE) was used to choose the top 10 features. With receiver operating characteristic (ROC) ratings of 91% and 97%, respectively, their binary classification findings were precise. However, the accuracy and ROC score of our suggested binary classification were 99.86% and 99.7%, respectively. Furthermore, all of the assaults in the sample were found using our intrusion detection technology. However, due to the usage of the recursive feature elimination-based feature selection technique, the key restriction is the amount of time required to create and train the model. Additionally, a sizable dataset must be used to assess the proposed model's efficacy [24].

The stratified k-fold cross-validation (SKCV) method was proposed by Prusty et al. to improve classification accuracy by removing redundant and weak features whose deletion had the least impact on the training error while retaining an independent and strong feature to improve the generalization performance of the model and address the overfitting problem (RFE). This method creates a model with the whole set of features before prioritizing them based on relevance. The model was then rebuilt with the lowest priority feature deleted, and the feature importance estimate was revised. However, developing and refining the model using the SKCV method takes a lot of time [25].

For the NIDS methodology, Kasongo and Sun combined the filter-based feature selection technique of the XGBoost algorithm with five classification algorithms: logistic regression (LR), k-nearest neighbors, artificial neural network (ANN), decision tree, and support vector machines. This study uses binary and multiclass classification using the UNSW-NB15 dataset. Multiclass classification performed poorly, with the maximum accuracy being just 82.66%, while binary classification using the k-nearest neighbor classifier did well, with an accuracy of 96.76%. However, a sizable dataset must be used to assess the suggested model's efficacy. The classification model is unaffected by the filter-based strategy of selecting characteristics as well [26].

The intrusion detection model proposed by Thaseen and Kumar uses a multi-class SVM classifier and a rank-based chi-square feature selection technique. The chi-squared test can be used to determine the deviation from the predicted distribution when the feature event is thought to be independent of the class value. A multi-class SVM is used to categorize the various sorts of attacks in the NSL-KDD dataset. Using the proposed model, 31 features were selected from a total of 41. The accuracy rate of the suggested system was 98%, while the false positive rate was 0.13% [27].

### III. PROPOSED METHODOLOGY

The methodology of this experimental study consists of five stages. First, the flow features were extracted using a CIC flow meter tool. In the second stage, dataset preprocessing was performed, and in the third stage, the best subset features were selected using the proposed PRFE technique. In the fourth step, a machine learning detection model is made. Finally, using the proposed PRFE technique and the ISOT-CID dataset, some supervised machine learning techniques are tested. Fig. 1 illustrates the five phases of the experimental study, which will be discussed in the remainder of this research.

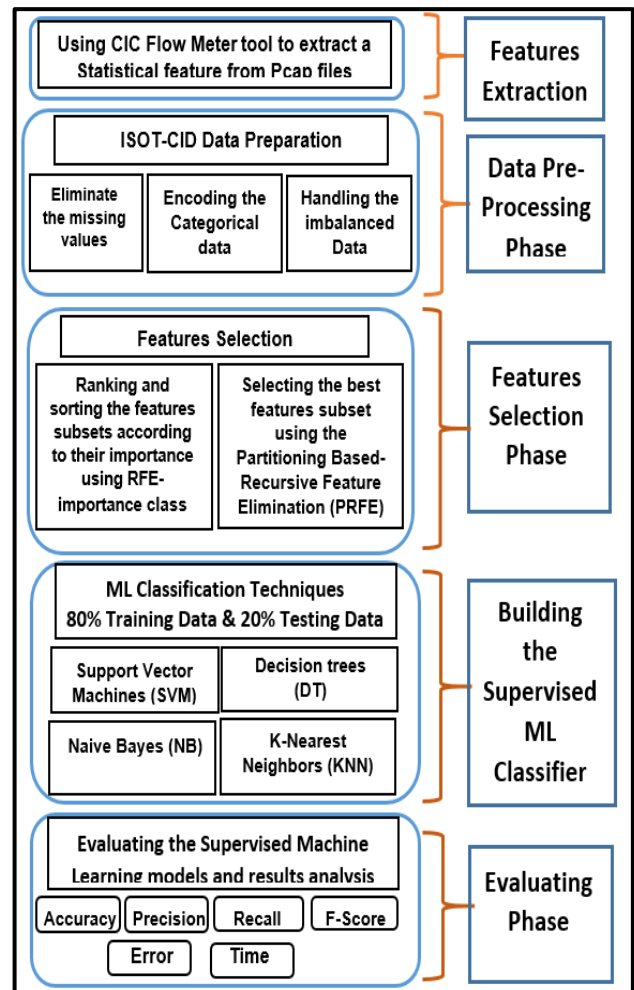


Fig. 1. Research methodology phases.



### A. Features Extraction

As shown in Fig. 2, the original ISOT-CID dataset consists of 12 attributes and 6,293,326 records [28]. In this study, a CIC flow meter was used to extract statistical and analytical features from the network flow. The CIC flow meter is a Java open-source tool that can generate and extract row attributes from huge Packet Capture (pcap) files and save the results in the comma-separated value (CSV) file format. Eighty-five network flow attributes were extracted from the ISOT-CID dataset using the CIC Flow Meter tool. The list of extracted features will be reduced to 80 after eliminating the five features that contained more than 90% missing data. Then, the proposed PRFE-based feature selection technique is used to choose the best, most important and most influential features of the learning process output [29]. Fig. 3 illustrates the number of features and records in the ISOT-CID dataset after the feature extraction stage.

### B. Dataset Preprocessing

The dataset used in the actual experiments in this study was ISOT-CID, which is considered the first huge, public, and labelled cloud intrusion detection dataset. The size of the ISOT-CID dataset is greater than 2.5 TB, and it consists of normal and malicious traffic activity collected from different cloud tiers, virtual machine hosts, and hypervisors. ISOT-CID data is collected in two phases and consists of different data formats, such as network traffic, CPU utilization, memory dumps, and event logs. The dataset consists of several types of attacks, including remote to local (R2L), input validation, backdoors, Denial of Service (DOS) and probing.

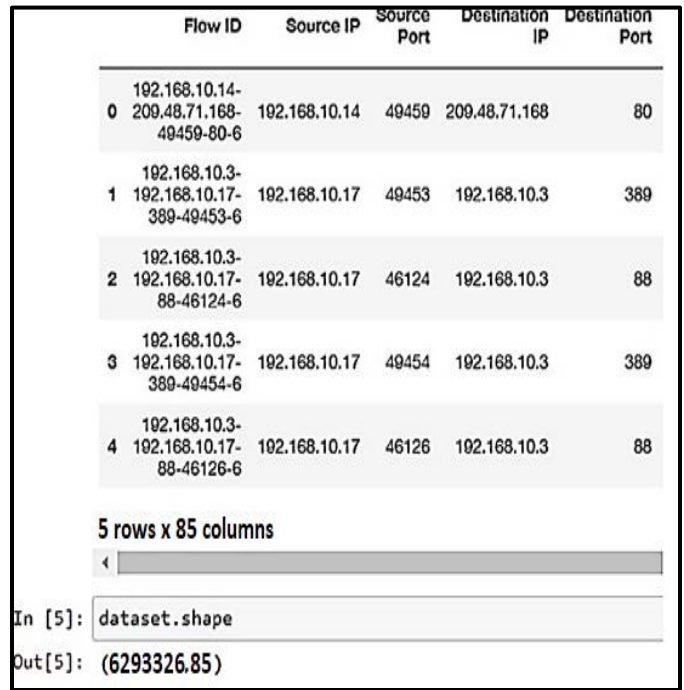


Fig. 3. ISOT-CID dataset after features extraction.

1) *Eliminate the missing values*: The dataset always needs to be reprocessed to remove duplicate and missing data before being used in the training phase of machine learning models, because relying on this dataset without processing would affect the quality of the learning results. A Python Jupyter Notebook was used to process and eliminate missing values in the ISOT-CID dataset. As shown in Fig. 4, some values in the flag, protocol, and fragment columns are missing.

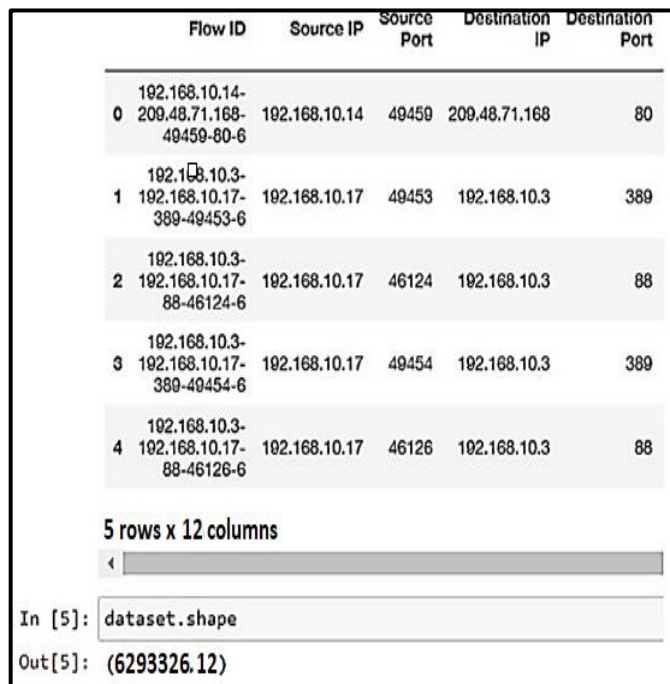


Fig. 2. ISOT-CID dataset before features extraction.

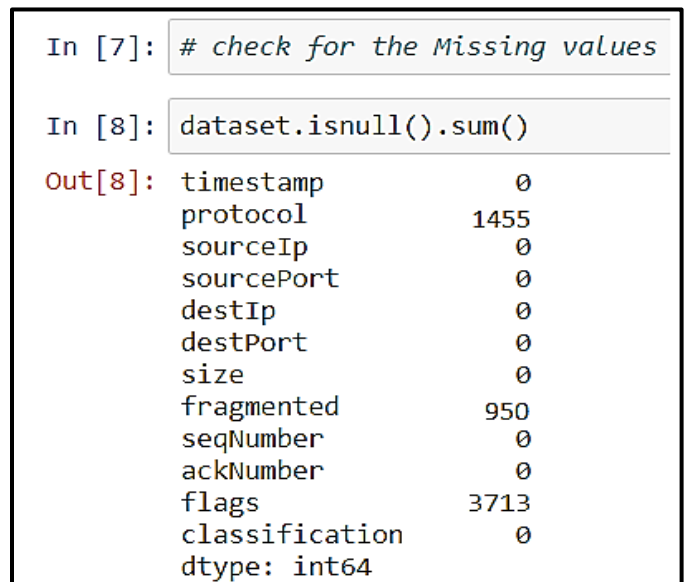


Fig. 4. ISOT-CID before handling missing values.

```
In [11]: dataset.dropna(inplace=True)
dataset.isnull().sum()

Out[11]: timestamp      0
protocol      0
sourceIp      0
sourcePort    0
destIp        0
destPort      0
size          0
fragmented    0
seqNumber     0
ackNumber     0
flags         0
classification 0
dtype: int64
```

Fig. 5. ISOT-CID after handling missing values.

Fig. 5 shows the ISOT-CID dataset after the missing values were taken care of by dropping them using the Python Jupyter Notebook. This improved the machine learning model's ability to classify and find malicious traffic.

2) *Label encoding for categorical data:* Dealing with machine learning algorithms to detect and classify malicious assaults in a big cloud environment requires encoding and converting some of the textual data that exists in the dataset into digital and numeric data to enhance and increase the level of accuracy of the learning results. As shown in Fig. 6, categorical attributes were converted to (0, 1) instead of (benign or malicious) attributes using the label encoding method.

```
In [11]: dummy = pd.get_dummies(dataset['classification'])

In [12]: dummy.head()

Out[12]:
```

	benign	malicious
0	1	0
1	1	0
2	1	0
3	1	0
4	1	0

Fig. 6. Encoding the categorical into numeric data.

3) *Handling the imbalanced-labelled data:* Class imbalance is a machine learning issue where the classes are not evenly represented in the data. This can cause issues while training machine learning models because the models may be biased towards the more prevalent class. The model will be more likely to pick up on and predict the majority class if

there are more samples of one class than the other. As a result, when the model is used to analyze data that is more evenly distributed, it may produce erroneous conclusions. Addressing unbalanced classification difficulties is a challenge in developing models with good performance. Fig. 7 illustrates the number of benign and malicious objects in the classification class before handling the imbalanced labelled data using the oversampling technique.

Synthetic Minority Oversampling Technique (SMOTE) is a method of oversampling, in which artificial samples are produced for the minority class. This method solves the overfitting problem caused by random oversampling. By interpolating nearby positive examples, we focused on the feature space to create new examples. Fig. 8 illustrates the number of benign and malicious objects in the classification class after handling the imbalanced labelled data.

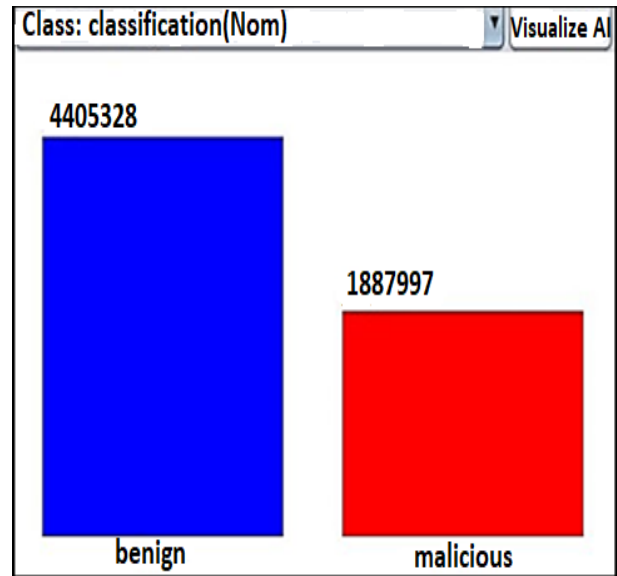


Fig. 7. Classification attribute before using SMOT.

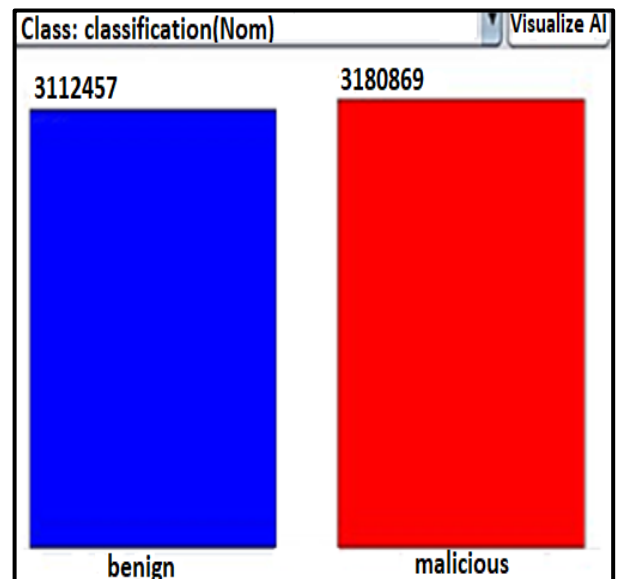


Fig. 8. Classification attribute after using SMOT.

### C. Proposed Features Selection Algorithm (PRFE)

Feature selection is a technique used to improve the accuracy of malicious attack classification by selecting the attributes that are the most important and significant to the outcome of the learning process and eliminating the least informative attributes. In this study, the ISOT-CID dataset was used to find the most useful attributes. To deal with labelled datasets, a supervised-based feature selection method was used [30]. Recursive feature elimination is a wrapper-based feature selection approach that assesses the significance of features using a machine-learning algorithm. All dataset features were used to train and fit the ML classifier model in the initial stages of recursive feature elimination, and the feature importance was calculated for each feature. The recursive feature elimination model is used repeatedly, with the least important features being thrown out and the most important ones being saved for the next round, until the best features are found. The partitioning-based recursive feature elimination (PRFE) technique was proposed in this study to improve the accuracy rate of classifying and detecting malicious attacks while reducing the complexity space and time required training the ML classifier models. Algorithm 1 illustrates the Partitioning-based Recursive Feature Elimination (PRFE) algorithm for selecting optimal features from the ISOT-CID dataset.

---

**Algorithm 1:** Partitioning based Recursive Feature Elimination (PRFE)

---

#Input

F: set of features where  $F = \{f_1, f_2 \dots f_n\}$

N: number of the required features

G: the number of groups where:  $1 < G \leq N$

$i = 1$

#Output

O: ordered ranked features.

R: ordered ranked groups.

Step 1: Train the model using all attributes

Step 2: Compute the model's accuracy

Step 3: Calculate and ranking the feature importance using the RFE-impotence class  $F_i^{\text{rank}}$  where  $i = 1 \dots N$  (N is the number of features).

Step 4: Divided the feature into equal number of groups (G), where numbers of features in each group are equal.

Note:  $G[i]$  contains number of features.

Step 5: Ranking and sorting the groups (G) in ascending order based on their features weight.

Step 6: Eliminating the lowest weighted group.

Step 7: Build ML classifier model and calculate model performance.

Step 8:  $i = i + 1$

Step 9: If  $i \neq G-1$

Repeat step 4 to step 7 until  $i = G-1$

End

---

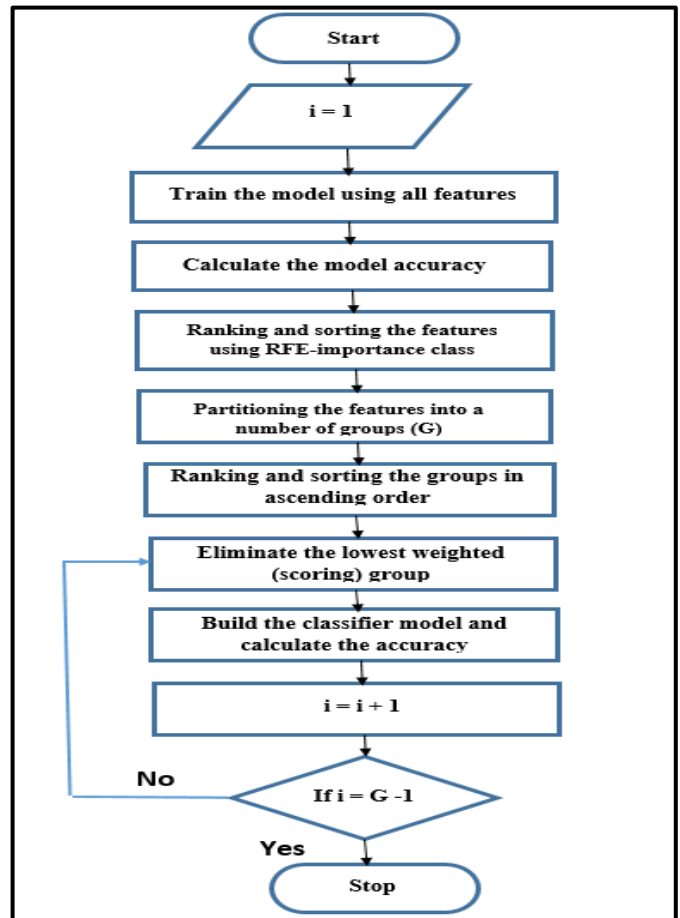


Fig. 9. Partitioning based Recursive Feature Elimination (PRFE) flowchart.

As shown in Fig. 9, the process of selecting optimal features from the ISOT-CID dataset using the PRFE technique occurs in two stages. In the first stage, the RFE importance class was used to rank and sort the feature subsets individually according to their importance and how strongly they were related to the outcomes of the learning process. In the second stage, the features are partitioned into groups, with an equal number of features in each group. For example, if we have 100 attributes, we can divide them into ten groups, each with ten attributes. Subsequently, the groups are ranked and sorted in ascending order, the lowest weighted group is eliminated in each iteration, and the training procedure for the remaining groups is repeated to obtain the best group of features. So, the number of tests went from 100 to 10, and the space and time needed to train the machine learning models became less complicated.

### D. Building the Machine Learning Classifier Models

For supervised learning, the ISOT-CID dataset was divided into training and testing sets. As a result, 80 % of the data were chosen at random and used to train machine-learning models, whereas the remaining 20% were utilized to evaluate the classifier's performance. Table I illustrates the statistics of the ISOT-CID Dataset used in this study.

TABLE I. STATISTICS OF THE ISOT-CID DATASET

Traffic Type	Total	Training 80%	Testing 20%
benign	3112457	2489965.6	622491.4
malicious	3180869	2544695.2	636173.8
Total	6293326	5034660.8	1258665.2

To evaluate the performance of the machine learning classifier models with different sets of selected features, accuracy, precision, recall, F-score, and error performance measurements were utilized. A confusion matrix was used to calculate the classifier performance indicators. "True positive" (TP) denotes benign instances that are correctly predicted, true negative (TN) denotes malicious instances that are correctly identified, false positive (FP) denotes malicious instances that are incorrectly assumed to be normal, and false negative (FN) denotes malicious instances that are incorrectly detected as normal [28]. Table II illustrates the five metrics that are commonly used to measure and evaluate the effectiveness of machine learning classification models.

TABLE II. PERFORMANCE METRICS FOR ML CLASSIFICATION MODELS

Metric	Formula	Interpretation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Overall performance of model
Precision	$\frac{TP}{TP+FP}$	How accurate the positive predictions are
Recall Sensitivity	$\frac{TN}{TN+FP}$	Coverage of actual positive sample
F1 score	$\frac{2TP}{2TP + FP+FN}$	Hybrid metric useful for unbalanced classes
Error Rate	$\frac{FP+FN}{TP+TN+FP+FN}$	the percentage of the classification that is done wrongly

E. Experiment Findings and Analysis

1) *The experiment setup:* Using the HP Z Book G3 workstation with Microsoft Windows 11 64-bit Enterprise edition and an Intel Core i7-6820HQ CPU @ 2.7GHz, 32GB RAM, the novel proposed PRFE approach was created using the Python version 3 code, which was implemented using the Jupyter Notebook platform and Anaconda virtual environment for Windows to execute Scikit-learn, NumPy, and Panda's libraries.

2) *The experiment findings:* Accuracy is one of the most important performance metrics in intrusion detection. The accuracy of the four supervised machine-learning classifiers using the proposed PRFE method outperformed the RFECV and RFE techniques in terms of overall performance. When PRFE-based selected features were used instead of RFECV- and RFE-based selected features, accuracy improved by approximately 0.75% and 2.25%, respectively. As shown in Fig. 10, with PRFE-based selected features, the Support

Vector Machine (SVM) classifier achieved the highest accuracy percentage of 99.25%. The k-nearest neighbor (KNN) classifier performed the worst in this trial. In general, the four machine-learning classifier models were more accurate after they used PRFE-based feature selection.

As shown in Fig. 11, the precision findings, which demonstrate the classifier's percentage of accurately identified instances, which is one of the key markers of excellent models. Classifiers trained with PRFE-based selected features outperformed those trained with RFECV and RFE-based selected features. When compared to other classifiers, the support vector machine (SVM) has the highest precision percentage of 98.80%. In contrast to prior results, the k-nearest neighbor (KNN) classifier has the lowest precision percentage in this trial, with a value of 96.50%. In general, the four machine-learning classifier models were more accurate when they used the PRFE-based feature selection method.

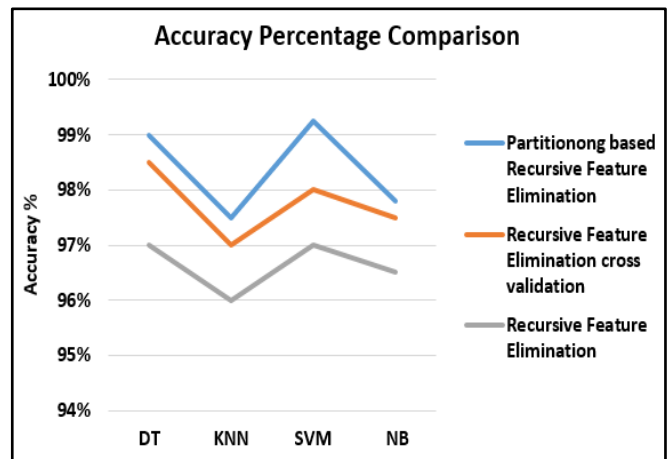


Fig. 10. Accuracy percentage comparison among PRFE, RFECV and RFE algorithm.

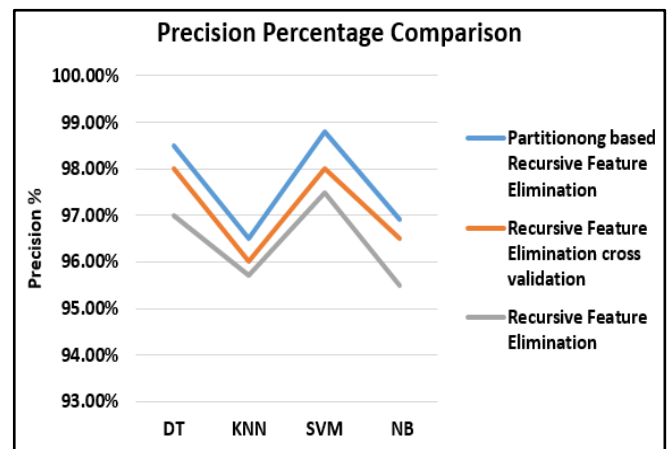


Fig. 11. Precision percentage comparison among PRFE, RFECV and RFE algorithm.

Fig. 12 shows the recall and sensitivity rates of the prediction models. Classifiers trained with PRFE-based selected features outperformed those trained with RFECV and RFE-based selected features. When compared to other classifiers, the decision tree classifier had the highest precision

percentage of 99%. In this experiment, naive Bayes exhibited the lowest precision percentage (97%). In general, the recall rate of the four machine learning classifier models got better when they used the PRFE-based feature selection method.

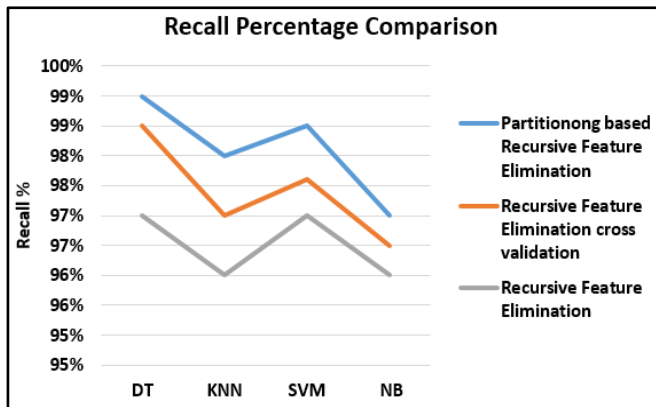


Fig. 12. Recall percentage comparison among PRFE, RFECV and RFE algorithm.

classifiers' f-scores performed better than when RFECV- and RFE-based selected features were used. With PRFE's chosen features, the support vector machine (SVM) classifier achieved the greatest f-score percentage of 99%. The k-nearest neighbor (KNN) classifier performed the worst in this trial with 97.5% (Table III).

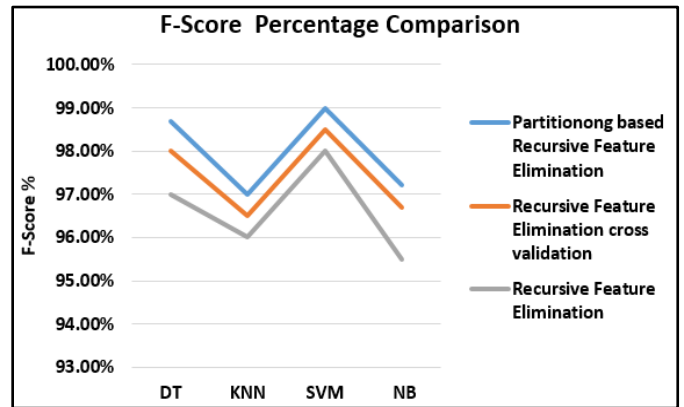


Fig. 13. F-Score percentage comparison among PRFE, RFECV and RFE algorithm.

Fig. 13 shows the f-score rate of the predictive models. In general, when PRFE-based selected features were used, the

TABLE III. OBSERVED ANALYSIS FOR DIFFERENT FEATURE SELECTION TECHNIQUES IN ISOT-CID USING ML CLASSIFIERS

Feature Selection Techniques	ML Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)	Error (%)	Time (m)
PRFE	DT	99%	98.50%	<u>99%</u>	98.70%	1%	2.33
	KNN	97.50%	96.50%	98.00%	97%	2.50%	1.55
	SVM	<u>99.25%</u>	<u>98.80%</u>	98.50%	<u>99%</u>	<u>0.75%</u>	<u>1.12</u>
	NB	97.80%	96.90%	97%	97.20%	1%	2.88
RFECV	DT	<u>98.50%</u>	98%	<u>98.50%</u>	98%	<u>1.50%</u>	6.88
	KNN	97%	96%	97%	96.50%	3%	5.65
	SVM	98%	<u>98%</u>	97.60%	<u>98.50%</u>	2%	<u>4.71</u>
	NB	97.50%	96.50%	97%	96.70%	2.50%	5.77
RFE	DT	<u>97%</u>	97%	97%	97%	<u>3%</u>	8.71
	KNN	95.50%	95.70%	96%	96.80%	4.50%	7.32
	SVM	96.50%	<u>97.50%</u>	<u>97%</u>	<u>98%</u>	3.50%	<u>6.71</u>
	NB	96.00%	96.00%	96.50%	96.00%	4.00%	7.77
Chi square	DT	97%	<u>97.50%</u>	<u>96.80%</u>	97%	3%	2.15
	KNN	96.50%	95%	94%	95.50%	3.50%	1.78
	SVM	<u>98.50%</u>	97%	95%	<u>98%</u>	<u>1.50%</u>	<u>1.55</u>
	NB	96%	95.80%	94.50%	96.40%	4%	2.45
Information Gain	DT	<u>98%</u>	<u>96.50%</u>	<u>97.50%</u>	96.50%	<u>2%</u>	1.85
	KNN	97%	96%	95%	96%	3%	1.25
	SVM	97.50%	96%	94.50%	<u>97%</u>	2.5%	<u>1.03</u>
	NB	96.50%	95%	94%	95.50%	3.50%	2.12
Backward Feature Elimination	DT	96.50%	<u>98%</u>	<u>97.80%</u>	<u>97.50%</u>	3.50%	5.44
	KNN	96%	95.50%	97%	96.50%	4%	4.65
	SVM	<u>97.50%</u>	96.50%	96%	96.70%	<u>2.50%</u>	<u>4.33</u>
	NB	96%	96.50%	95.40%	95%	4%	5.12

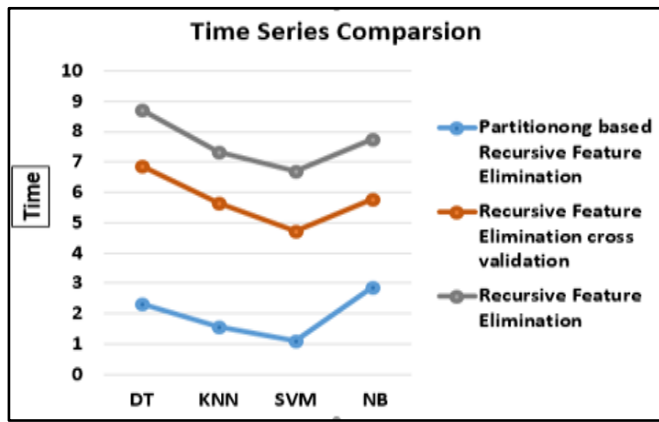


Fig. 14. Time series comparison among PRFE, RFECV and RFE algorithm.

Fig. 14 indicates a decrease in the time required to train the model using the proposed PRFE method compared to RFECV and RFE techniques.

#### IV. DISCUSSION OF RESULTS

The ISOT-CID dataset was used for the evaluation experiments, which were carried out using Python programming and the Jupyter Notebook. In the evaluation experiment, the proposed PRFE was compared with RFE- and RFECV-based feature selection techniques. The comparison with the other approaches is defined in terms of accuracy, precision, recall, F-score, and time series. The evaluation trial demonstrates PRFE's advantages over competitors. Therefore, in this section, we will discuss how it stacks up against more modern research methods.

The traditional methods for selecting the most important features in the previous studies were characterized by simplicity and the low time required training machine learning models, while suffering from a low level of accuracy in identifying and detecting malicious attacks. Modern methods focus on raising the level of accuracy while neglecting the time required training machine learning models. Therefore, the proposed PRFE technique enhances the accuracy rate by combining supervised machine learning classifier models with partitioning-based recursive feature elimination techniques. This led to an increase in the level of accuracy in identifying and detecting malicious attacks to 99.25, while simultaneously reducing the time required to train automated training models to 1.2 minutes. Because this study deals with the cloud-based intrusion detection system (CIDS), the proposed PRFE technique was evaluated using the ISOT-CID dataset, which is considered one of the first public datasets of its kind collected from a production cloud environment. Most previous studies lacked datasets from a real cloud computing environment that were available to the public. This made it hard to make and test realistic detection models.

#### V. CONCLUSION AND FUTURE WORKS

Feature engineering techniques such as data preprocessing, feature extraction, and feature selection should be used to reduce the dimensionality of the input features, improve model performance, and shorten model computational time. Choosing the most relevant and influential characteristics has

traditionally affected the power and predictability of the final classifier model. The most pertinent and effective features were chosen from large datasets using a variety of feature selection techniques, including chi-square, information gain, backward feature elimination, and recursive feature elimination. In this study, a PRFE-based feature selection method was developed to classify and select the optimal feature subset from ISOT-CID, which is considered one of the largest public intrusion detection datasets. The best feature subset was selected by partitioning the features into groups with an equal number of features in each group and eliminating the lowest-weighted group in each iteration. The PRFE method improved accuracy, precision, recall, and F-score rate while cutting training time by ignoring one group instead of removing one feature at each iteration.

In this experimental study, the proposed PRFE-based feature selection technique was first compared with recursive feature elimination (RFE) and recursive feature elimination with cross-validation (RFECV) techniques. The results showed that the proposed PRFE technique improved the accuracy, precision, recall, and F-score percentage with the four common machine learning classifier models compared to the RFE and RFECV techniques. Second, a few popular filter- and wrapper-based feature selection methods, including chi-square, information gain, and backward feature removal, are compared to the proposed PRFE-based feature selection strategy. The results of the experiments show that when the PRFE method is used with four popular machine learning classifier models, the accuracy, precision, recall, and F-score percentage are all higher than when the chi-square, information gain, and backward feature elimination strategies are used. In future work, a new unsupervised deep learning algorithm for detecting zero-day attacks will be proposed. This algorithm will use different neural network topologies, such as fully connected, recurrent, and temporal convolutional models, to reduce the number of false alarms while maintaining the accuracy of detecting and classifying malicious network attacks.

#### REFERENCES

- [1] Butt, U., Mehmood, M., Shah, S., Amin, R., Shaukat, M., & Raza, S. et al. (2020). A Review of Machine Learning Algorithms for Cloud Computing Security. *Electronics*, 9(9), 1379. <https://doi.org/10.3390/electronics9091379>.
- [2] Xu, Z., Liu, Y., Yen, N., Mei, L., Luo, X., Wei, X., & Hu, C. (2020). Crowdsourcing Based Description of Urban Emergency Events Using social media Big Data. *IEEE Transactions on Cloud Computing*, 8(2), 387-397. <https://doi.org/10.1109/tcc.2016.2517638>.
- [3] Sarker, I., Kayes, A., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big Data*, 7(1). <https://doi.org/10.1186/s40537-020-00318-5>.
- [4] Megantara, A., & Ahmad, T. (2021). A hybrid machine learning method for increasing the performance of network intrusion detection systems. *Journal of Big Data*, 8(1). <https://doi.org/10.1186/s40537-021-00531-w>.
- [5] Mebawodu, J., Alowolodu, O., Mebawodu, J., & Adetunmbi, A. (2020). Network intrusion detection system using supervised learning paradigm. *Scientific African*, 9, e00497. <https://doi.org/10.1016/j.sciaf.2020.e00497>.
- [6] Jadhav, A., & Pellakuri, V. (2021). Highly accurate and efficient two phase-intrusion detection system (TP-IDS) using distributed processing of HADOOP and machine learning techniques. *Journal of Big Data*, 8(1). <https://doi.org/10.1186/s40537-021-00521-y>.

- [7] Taylor, O., & Ezekiel, P. (2021). Anomaly-Based Intrusion Detection/Prevention System using Deep Reinforcement Learning Algorithm. *IJARCCCE*, 10(1). <https://doi.org/10.17148/ijarccce.2021.10114>.
- [8] Panda, M., Mousa, A., & Hassanien, A. (2021). Developing an Efficient Feature Engineering and Machine Learning Model for Detecting IoT-Botnet Cyber Attacks. *IEEE Access*, 9, 91038-91052. <https://doi.org/10.1109/access.2021.3092054>.
- [9] Krishnaveni, S., Sivamohan, S., Sridhar, S., & Prabakaran, S. (2021). Efficient feature selection and classification through ensemble method for network intrusion detection on cloud computing. *Cluster Computing*, 24(3), 1761-1779. <https://doi.org/10.1007/s10586-020-03222-y>.
- [10] Umar, M., Chen, Z., & Liu, Y. (2021). A Hybrid Intrusion Detection with Decision Tree for Feature Selection. *Information & Security: An International Journal*. <https://doi.org/10.11610/isij.4901>.
- [11] Kadhum, M., Manaseer, S., & Dalhoum, A. (2021). Evaluation Feature Selection Technique on Classification by Using Evolutionary ELM Wrapper Method with Features Priorities. *Journal of Advances in Information Technology*, 12(1), 21-28. <https://doi.org/10.12720/jait.12.1.21-28>.
- [12] Viharos, Z., Kis, K., Fodor, Á., & Büki, M. (2021). Adaptive, Hybrid Feature Selection (AHFS). *Pattern Recognition*, 116, 107932. <https://doi.org/10.1016/j.patcog.2021.107932>.
- [13] Xiao, W., Ji, P., & Hu, J. (2021). RnkHEU: A Hybrid Feature Selection Method for Predicting Students' Performance. *Scientific Programming*, 2021, 1-16. <https://doi.org/10.1155/2021/1670593>.
- [14] Das, A., -, P., & S, S. (2022). Anomaly-based Network Intrusion Detection using Ensemble Machine Learning Approach. *International Journal of Advanced Computer Science and Applications*, 13(2). <https://doi.org/10.14569/ijacsa.2022.0130275>.
- [15] Dhanda, N., Datta, S., & Dhanda, M. (2019). Machine Learning Algorithms. *Computational Intelligence In The Internet of Things*, 210-233. <https://doi.org/10.4018/978-1-5225-7955-7.ch009>.
- [16] Gaydamaka, K., & Belonogova, A. (2022). Applying Unsupervised Machine Learning Algorithms to Ensure Requirements Consistency. *Programmnaya Ingeneria*, 13(4), 187-199. <https://doi.org/10.17587/prin.13.187-199>.
- [17] Peng, K., Leung, V., Zheng, L., Wang, S., Huang, C., & Lin, T. (2018). Intrusion Detection System Based on Decision Tree over Big Data in Fog Environment. *Wireless Communications and Mobile Computing*, 2018, 1-10. <https://doi.org/10.1155/2018/4680867>.
- [18] Belouch, M., El Hadaj, S., & Idhammad, M. (2018). Performance evaluation of intrusion detection based on machine learning using Apache Spark. *Procedia Computer Science*, 127, 1-6. <https://doi.org/10.1016/j.procs.2018.01.091>.
- [19] Belavagi, M., & Muniyal, B. (2016). Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection. *Procedia Computer Science*, 89, 117-123. <https://doi.org/10.1016/j.procs.2016.06.016>.
- [20] Azeroual, O., & Nikiforova, A. (2022). Apache Spark and MLlib-Based Intrusion Detection System or How the Big Data Technologies Can Secure the Data. *Information*, 13(2), 58. <https://doi.org/10.3390/info13020058>.
- [21] Souhail et. al., M. (2019). Network Based Intrusion Detection Using the UNSW-NB15 Dataset. *International Journal of Computing and Digital Systems*, 8(5), 477-487. <https://doi.org/10.12785/ijcds/080505>.
- [22] Alshammari, A., & Aldribi, A. (2021). Apply machine learning techniques to detect malicious network traffic in cloud computing. *Journal of Big Data*, 8(1). <https://doi.org/10.1186/s40537-021-00475-1>.
- [23] Aslahi-Shahri, B., Rahmani, R., Chizari, M., Maralani, A., Eslami, M., Golkar, M., & Ebrahimi, A. (2015). A hybrid method consisting of GA and SVM for intrusion detection system. *Neural Computing and Applications*, 27(6), 1669-1676. <https://doi.org/10.1007/s00521-015-1964-2>.
- [24] Mohammed, B., & Gbashi, E. (2021). Intrusion Detection System for NSL-KDD Dataset Based on Deep Learning and Recursive Feature Elimination. *Engineering and Technology Journal*, 39(7), 1069-1079. <https://doi.org/10.30684/etj.v39i7.1695>.
- [25] Prusty, S., Patnaik, S., & Dash, S. (2022). SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer. *Frontiers in Nanotechnology*, 4. <https://doi.org/10.3389/fnano.2022.972421>.
- [26] Kasongo, S., & Sun, Y. (2020). Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset. *Journal of Big Data*, 7(1). <https://doi.org/10.1186/s40537-020-00379-6>.
- [27] Sumaiya Thaseen, I., & Aswani Kumar, C. (2017). Intrusion detection model using fusion of chi-square feature selection and multi class SVM. *Journal of King Saud University - Computer and Information Sciences*, 29(4), 462-472. <https://doi.org/10.1016/j.jksuci.2015.12.004>.
- [28] Aldribi, A., Traoré, I., Moa, B., & Nwamuo, O. (2020). Hypervisor-based cloud intrusion detection through online multivariate statistical change tracking. *Computers & Security*, 88, 101646. <https://doi.org/10.1016/j.cose.2019.101646>.
- [29] Habibi Lashkari, A., Draper Gil, G., Mamun, M., & Ghorbani, A. (2017). Characterization of Tor Traffic using Time based Features. *Proceedings of the 3Rd International Conference on Information Systems Security and Privacy*. <https://doi.org/10.5220/0006105602530262>.
- [30] Alabdulwahab, S., & Moon, B. (2020). Feature Selection Methods Simultaneously Improve the Detection Accuracy and Model Building Time of Machine Learning Classifiers. *Symmetry*, 12(9), 1424. <https://doi.org/10.3390/sym12091424>.

# Parameter Extraction and Performance Analysis of 3D Surface Reconstruction Techniques

Richha Sharma<sup>1</sup>, Pawanesh Abrol<sup>2</sup>  
Dept. of CS & IT, University of Jammu, Jammu, India<sup>1,2</sup>

**Abstract**—Digital image-based 3D surface reconstruction is a streamlined and proper means of studying the features of the object being modelled. The generation of true 3D content is a very crucial step in any 3D system. A methodology to reconstruct a 3D surface of objects from a set of digital images is presented in this paper. It is simple, robust, and can be freely used for the construction of 3D surfaces from images. Digital images are taken as input to generate sparse and dense point clouds in 3D space from the detected and matched features. Poisson Surface, Ball Pivoting, and Alpha shape reconstruction algorithms have been used to reconstruct photo-realistic surfaces. Various parameters of these algorithms that are critical to the quality of reconstruction are identified and the effect of these parameters with varying values is studied. The results presented in this study will give readers an insight into the behaviour of various algorithmic parameters with computation time and fineness of reconstruction.

**Keywords**—3D reconstruction; point cloud; feature detection; feature matching

## I. INTRODUCTION

The 3D reconstruction process can generate the photo-realistic 3D surfaces of objects using their digital images to generate their sparse and dense 3D point clouds. The generated point cloud (PC) data is to be transformed into a substantial digital representation of the scanned objects to enable its use in numerous application fields including Computer-Aided Design, Medical Imaging, Reverse Engineering, Virtual Reality, and Architectural Modeling [1]. Since object surfaces in point clouds are to be reconstructed, various reconstruction approaches have been proposed over years to fast-track the research in this area. Earlier, 3D reconstruction was mainly applied in the field of Industrial and Architectural design, Animation industry, Medical Modeling, Robotics, etc. but with the advent of Machine Learning and 3D vision technology, the application range has widened to other fields like autonomous driving, remote sensing, 3D printing, and even online shopping. The sample points can be generated from the laser scanner, photogrammetry technique, or some mathematical function. In most cases, the sample points will elucidate the shape and topology of the object's surface. Algorithms for surface reconstruction (SR) are applied to the sample points, and to obtain a 3D model of the object [2]. 3D reconstruction can usually be done in three manners 1) Using 3D modeling software like AutoCAD, FreeCAD, 3D MAX, etc. 2) Using 3D scanners to obtain models in real-time. 3) Using digital images with the help of RGB-D cameras. Close-range photogrammetry has been dealing with manual or

programmed picture measurements for accurate 3D modelling. Although many applications utilize 3D scanners as a regular source of data input, image-based modelling is still the most thorough, affordable, portable, adaptable, and widely-used method. It is a low-priced and easy way to obtain depth images. Point cloud obtained from RGB-D data is of middle density with the advantages of being cheap and flexible along with the color information. It is limited to close-range objects with moderate accuracy in applications like Indoor reconstruction, object tracking, human pose recognition; etc. Laser-scanner-based 3D reconstruction of an object is quick and accurate and provides realistic textured 3D surfaces of a scene or object being studied. Because of the high cost and less availability, researchers have been inspired to use photogrammetry to produce realistic 3D surfaces from an object's image. It is used for mapping and modeling common indoor household objects. With the ease of availability of digital cameras and computers in homes and organizations, the processing cost of image-based reconstruction has diminished. The work in this paper evaluates the efficacy of various surface reconstruction techniques. It proposed a framework for reconstructing 3D surfaces that can generate a 3D model of typical indoor objects from their digital photos. Correspondence matching is done over the image sets to obtain a dense set of points that are projected to 3D space. Furthermore, by applying SR techniques to the resulting point cloud, a photo-realistic surface is formed.

The paper is structured as: Section II presents a summary of surface reconstruction algorithms and work related to 3D reconstruction. Section III presents the framework of the used methodology along with a description of the selected algorithms used for reconstruction. Section IV provides a discussion of the experiment with the obtained results. The conclusion and future outline of the work is given in Section V.

## II. RELATED WORK

Reconstructing a 3D surface from an order of digital images comprising a scene/object is a demanding and significant exercise in computer vision. Over the years, authors have employed several SR approaches to create a precise and photorealistic 3D surface from an image or group of images depicting a variety of objects. 3D reconstruction from images (photogrammetry), is a long-standing technique whose potential could only be seen after the development of computers and digital photographs and is widely used in entertainment to create scenarios for games, acquire human expressions, scanning products for commercials, etc.



TABLE I. DESCRIPTION OF SURFACE RECONSTRUCTION ALGORITHMS

Authors	Year	Algorithm	Description
Edelsbrunner and Mucke [3]	1994	Alpha shape algorithm	It is a heuristic approach and performs well in uniform sampling and has the complexity of $O(n^2)$
Bernardini <i>et al</i> [4]	1999	Ball-Pivoting algorithm	Forms a 3D mesh by joining the triangles created by rolling the ball through points. It is data-driven and is sensitive to noise and is not expensive in terms of time and memory and results are of favorable quality.
Amenta <i>et al.</i> [5]	2000	The Crust algorithm	Delaunay Triangulation is estimated employing the Voronoi diagrams over the point cloud forming a smooth surface from unstructured points. It is robust and does not require a hole-filling mechanism. It has the complexity of $O(n \log n)$
Amenta <i>et al.</i> [6]	2001	Power Crust algorithm	It gives approximated polygon surface with the amalgamation of median balls. The algorithm is robust and no hole-filling structure is needed in reconstruction as the factual geometry of the object is captured.
Amenta <i>et al.</i> [7]	2001	Cocone algorithm	Delaunay-based algorithm for surface reconstruction with added post-filtering restrictions. It produces an incomplete surface with a low-density point cloud.
Dey and Goswami [8]	2003	Tight Cocone algorithm	The reconstructed surface uses the 'In' marked computed Cocone tetrahedral triangles. This method is not robust to a noisy and sparse point cloud and has a time complexity of $O(n^2)$ .
Dey and Goswami [9]	2004	Robust Cocone algorithm	It can reconstruct surfaces from noisy points and is also considered a labeling algorithm as it marks cocone triangles as "in" and "out".
Kazhdan <i>et al.</i> [10]	2006	Poisson Surface Reconstruction	The watertight surface is recreated with Marching Cubes (MC) algorithm and stored in an octree. The algorithm is resilient to noise.
Kazhdan <i>et al.</i> [11]	2013	Screened Poisson Surface Reconstruction	It incorporates point constraints in solving the screened Poisson equation to compose the solved indicator function more accurately.

### A. Surface Reconstruction Techniques

From input data, which is classically obtained from 3D scanners or photogrammetry as scattered points in 3D space, the process of SR creates a 3D surface of the object [12]. The surface is then often recreated in the triangulated form (points joined in the form of triangles with sharing edges and/or vertices), depending upon the type of point data employed [13] making a visible, 3D model of a real object. In the field of computer graphics, numerous surface reconstruction techniques have been developed and used. In this paper, a few of the well-known surface reconstruction algorithms are briefly discussed and presented in Table I. A brief description of a few of the algorithms is given in hierarchical order in terms of the year in which they were introduced.

The primary input of the SR process is a digital image set of the object being reconstructed. This is done by detecting features from the image that will help in the process of reconstruction. These feature points are searched for correspondence; projection of the same 3D point of the object across the images [14]. Two feature detection algorithms have been used. SURF (Speeded up Robust Features) focuses on blob-like structures in the image [15]. It is fast, robust, and invariant to rotation and scale which makes it a good feature detection algorithm [16]. Min Eigen Features detect the corners (locations with steep intensity variations in any direction) in an image. Feature matching, generally known as image matching is performed in many computer vision applications like image registration, object identification, and object recognition. It consists of detecting a set of interest points each associated with image descriptors from image data. After the extraction of features from the image set, corresponding feature points are matched. Matching shows that features are from the corresponding locations from completely different images.

There are certain issues and constraints while dealing with point data. The unstructured-ness of data makes it difficult to

implement and reconstruct the surface because of the non-uniformity of input data. Dataset size is another major constraint and can lead to a tradeoff between accuracy and performance. Performance has been always a bottleneck in all computer-based applications in terms of time, memory efficiency accuracy, and completeness.

### III. METHODOLOGY

A digital image sequence of the scene or item being recreated serves as the primary input to the SR process. Fig. 2 demonstrates the process used to recreate an object's 3D surface from a series of digital images. A set of distinguished feature points are found in the image sequence and used for correspondence matching across the photos to begin the 3D reconstruction process. The exhaustive matching of all image points is considerably reduced by feature detection, making it computationally cheap. The robust feature points are also easily recognizable and image transformation-invariant. Consequently, by identifying feature points in the photos, the computation time for correspondence matching is significantly decreased.

The dataset used in the work comprises images of irregular indoor household objects to be reconstructed. It consists of first-hand image data comprising 4 image sets taken from an android phone having a (16+20) MP dual camera placed at a distance of 25 cm from the object. Images were preprocessed to a standard form for further use and Table II shows the parameter values of preprocessed images. Fig. 1 presents the image dataset used as input in this paper for the implementation purpose and the attributes of the images are given in Table II. Image features were detected and extracted from all the images containing the object of interest using the SURF (Speeded Up Robust Features) algorithm. The SURF-identified feature points are readily distinct and invariable to image transformations [17] and hence are then used for correspondence matching.

TABLE II. PARAMETER VALUES OF IMAGES USED FOR IMPLEMENTATION

Parameter	Value
Size of image	1520 x 2048
Min-Quality	0.1
Confidence	99.90
Roi (x, y, width, height)	[30, 30, 2018, 1490]



Fig. 1. The input image dataset.

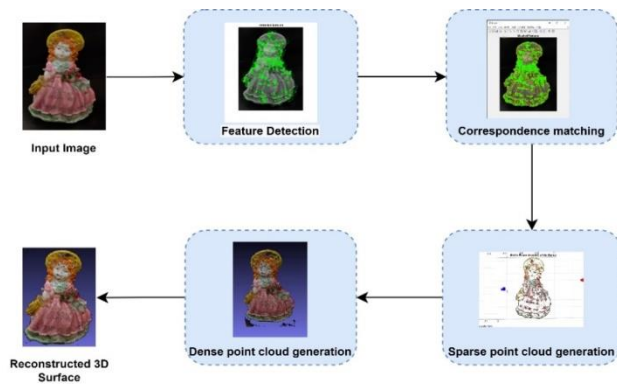


Fig. 2. Methodology of 3D surface reconstruction.

An approximate-nearest-neighbor (ANN) search discovers the matched point pairs. An image pair of the object with enough overlap has been selected for the purpose of matching. The features of the first image were inserted in the k-d tree which acts as queries to the feature points of other images of the pair. The k-d tree has been used for the efficient search of K-nearest neighbors of the feature points. The accuracy of the paired points is then tested using a RANSAC-based estimation. Matches that do not satisfy the basic matrix equation are discarded as non-fits.

SFM has been applied to build sparse point clouds (SPC) from huge image sets. In SFM, correspondence matching finds the matched points from the image set that are triangulated to obtain the 3D location of these points. The obtained point clouds are not dense and do not comprehend the object's geometry or specifics of the scene under reconstruction, hence are inadequate for 3D modelling [18]. Thus, dense correspondence matching must be done to build dense point clouds (DPC) from sparse points of the image sets [19], precisely unfolding the specifics of the scene/object being reconstructed and as shown in Fig. 3.

The pronounced methodology for 3D reconstruction used in this paper is grounded on several well-known algorithms in the fields of correspondence matching, SFM, SR, and so on. Rather than delving into the thorough mathematics of these

techniques, the study focused on the application of existing techniques to create an inexpensive, robust, and simple answer to the problem of 3D reconstruction. Because of its robustness, swiftness, and capability to reconstruct surfaces with irregular geometry, the approach described in this paper is easily applicable to industry. The photogrammetry-based reconstruction method conversed here can efficiently substitute expensive laser scanners as prime 3D reconstruction methods in numerous application areas due to its little workforce requirement and reduced functioning cost. Three surface reconstruction algorithms were tested on a PC generated from digital images. The performance of these reconstruction algorithms is evaluated in terms of time, resource usage, reconstruction quality, and so on. The following section provides brief overviews of these algorithms.



Fig. 3. 3D point cloud of object Image (row1 sparse point cloud, row 2 dense point cloud).

#### A. Poisson Surface Reconstruction (PSR)

PSR intends to create a 3D mesh from a DPC by minimalizing the difference between the surface normal directions of the recreated surface and the 3D points [10] (Fig. 4). The surface normal of each point's surface for the input point cloud is found by computing the eigenvectors over the k-nearest neighbor of each point, and an octree of pre-defined depth is used for storing the reconstructed surface. A 3D indicator function  $x$  is defined as having the value of 1 inside and 0 outside the surface.  $\nabla x$  (grad of the indicator function) is a vector field comprising non-zero values at the points near the surface where  $x$  can be varied.

For the points where  $\nabla x$  is not zero, the value of  $\nabla x$  equates to the surface normal vector of corresponding points and is stated as:

$$\nabla x = V \quad (1)$$

On both sides of the equation, the divergence operator is applied to modify the equation into a standard Poisson problem and can be articulated mathematically as:

$$\Delta x \equiv \nabla \cdot \nabla x = \nabla \cdot V \quad (2)$$

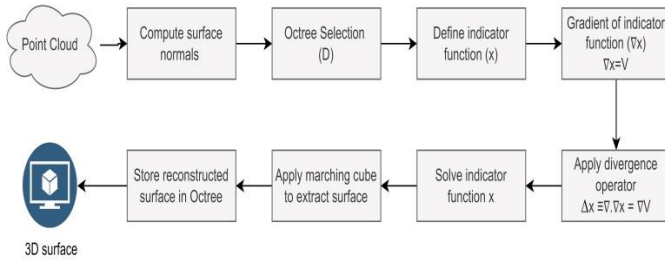


Fig. 4. Flow chart of Poisson surface reconstruction.

The indicator function ( $x$ ) is represented in 3D space for solving the Poisson equation. Octree represents the indicator function and its leaf nodes store  $x$  at various points over the reconstructed surface. The MC algorithm extracts the surface from the indicator function. The algorithm works by marching through the points in a 3D point cloud following the indicator function having a value of 1 close to the surface and 0 afar from the surface. By interpolation of points among the vertices of the cube during the traversal of the octree, a triangulated 3D mesh is generated and stored in the octree. The quality of reconstruction, computation time, etc. of PSR depends upon several parameters like Octree depth ( $D$ ), Samples-per-node ( $S_n$ ), and Surface offsetting ( $S_{off}$ ). An octree with depth  $D$  generates a 3D mesh having resolution ( $2^D \times 2^D \times 2^D$ ) that rises with the rise in  $D$  value and hence, the memory usage also rises severely.  $S_n$  indicates the least number of points that are allocated to individual leaf nodes of octree by the MC algorithm. If the data is noisy, a greater number of points are allocated as the surface is interpolated with the entire data points.  $S_{off}$  specifies a correction value of the threshold for the reconstructed surface.  $S_{off}=1$ , no correction required,  $S_{off} < 1$  signifies internal offsetting, and for external offsetting  $S_{off} > 1$ .

### B. Alpha-shape

It contains points, edges, triangles, and tetrahedrons. A subgroup of the triangles from the Delaunay triangulation is carefully selected to obtain a surface. The initial triangle with the property of minimum area is considered as the 'seed' which is further used to propagate the whole surface [3]. A subset of the 3D Delaunay triangulation of  $P$  is the  $\alpha$ -complex of a set of points ( $P$ ) for a given value of the parameter, with the  $\alpha$ -shape serving as its underlying space. As the value of  $\alpha$  varies, different  $\alpha$ -shapes can be generated. For  $\alpha = 0$  the  $\alpha$ -shape is  $P$  itself and when  $\alpha = \infty$ , the resultant shape is the convex hull of the point set. Changing the value of  $\alpha$  from 0 to  $\infty$ , different  $\alpha$ -shapes can be attained.

### C. Ball Pivoting Algorithm (BPA)

It is an effective surface reconstruction technique in which a ball of the pre-defined radius is rolled across the point cloud [4]. Triangular inter-connected 3D mesh is generated as the ball moves through the cloud, connecting the 3D points. The process is repeated until all the points are linked to a triangle. The algorithm is divided into two steps. The primary step is to find a seed triangle, and the second step is to expand the seed triangle to create a 3D surface, as shown in Fig. 5 and 6. For the input points, the surface normal for all points is computed and a point ( $p$ ) is selected.

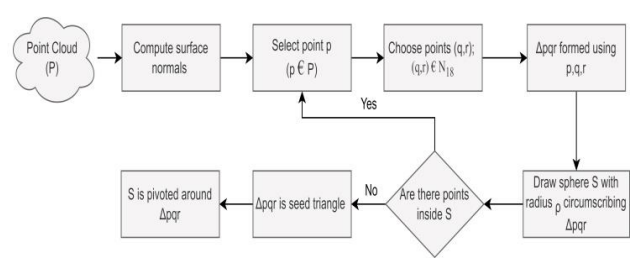


Fig. 5. Flow chart of ball pivoting algorithm (Step I).

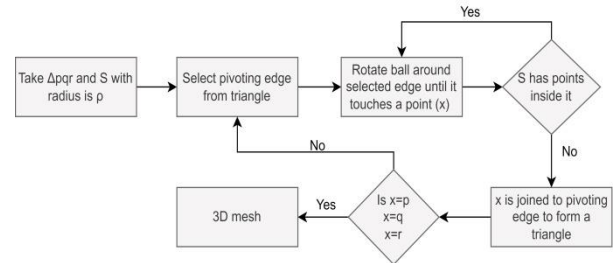


Fig. 6. Flow chart of ball pivoting algorithm (Step II).

A triangle is formed by picking two nearest neighbors' triangle if the sphere ( $S$ ) has no point inside it else, a new point is chosen from the point cloud. The seed triangle is extended by rolling the ball that is pivoted to the chosen edge of  $\Delta pqr$  to find an alternative point in the point cloud. Another triangle is formed by joining the new-found point to the pivoting edge. The process is repeated until all the points have been navigated and linked by triangles to form a 3D mesh, as shown in Fig. 7. ( $q,r$ ) of  $p$  such that  $(q,r) \in N18$  resulting in  $\Delta pqr$  and its circumsphere is searched for points.  $\Delta pqr$  is the seed triangle.

BPA is sensitive to variations in ball radius ( $\rho$ ) which acts as a critical parameter in defining the number of reconstructed faces. A small radius causes the model to be vulnerable to input noise while a high value of radius can result in missing details creating holes in the generated surface. Another parameter is the angle threshold which is the maximum permissible angle between the active and the new edge constructed by rolling the ball. The rolling of the ball is stopped for the region where the angle threshold is achieved. An upsurge in threshold angle increases the computation time. The clustering radius is the lowest permitted distance between a recently added point and the active edge point. If there is less space separating two locations than the clustering radius, the two are combined. This is done to prevent processing DPC from consuming too much memory as a result of creating too many little triangular meshes.

The methodology defined and implemented in this study is based on well-known algorithms in the area of correspondence matching, structure-from-motion, surface reconstruction, etc. Instead of delving into the intricate mathematics involved in creating these algorithms, the focus here is on using these techniques to create a practical, reliable, and straightforward solution to the 3D reconstruction task. The photogrammetry-based 3D reconstruction strategy detailed here can successfully replace the pricey laser scanners as the principal

3D mapping method in industries due to its minimal labor demand and nearly zero running cost [20].

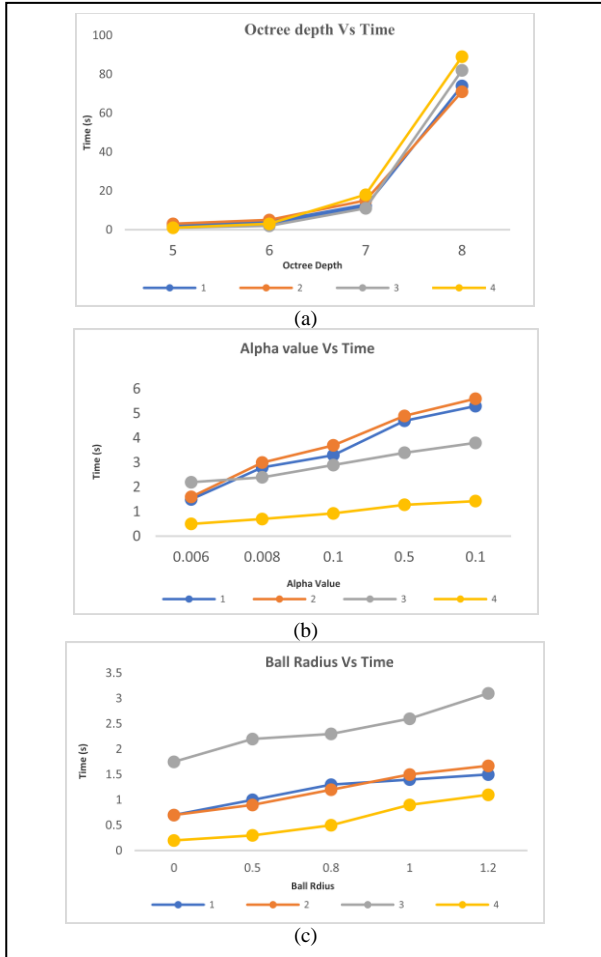


Fig. 7. Behavior of different parameters for selected algorithms with respect to time (a) PSR (b) Alpha Shape (c) BPA.

#### IV. EXPERIMENTAL RESULTS

The experimental work has been performed on Intel Core (TM) i7 CPU with 20 GB RAM accompanied by NVIDIA GeForce 1050Ti. MATLAB 2021 on windows 10 has been used for implementation. To analyze how well the 3D reconstruction algorithms performed, some point clouds were produced from digital image sets of various objects. To create a realistic 3D surface of the acquired item, PSR, Alpha shape, and BPA algorithms have been applied to the point clouds. The topology surface, the number of faces and the amount of time the algorithm took to compute each 3D surface are all taken into consideration in the analysis. The parameters affecting the reconstruction process were inspected in detail, by altering the control parameters of respective algorithms in terms of the execution time and reconstruction fineness. Fig. 6 shows the graphs of selected algorithmic parameters with respect to time. The primary influencing parameter in PSR is Octree Depth (D), and execution time grows with depth as the resolution of the 3D mesh increases by  $2^D \times 2^D \times 2^D$ . It has been found that an increase in alpha value causes an increase in execution time in the Alpha shape technique, which is a

critical parameter for the quality of the rebuilt surface. The ball radius is an important parameter in BPA, and it has been found that as grows in value, surface reconstruction takes longer to complete. The performance of selected reconstruction algorithms has been evaluated with the same input images and is compared in terms of how fine is the reconstruction. The final triangulated surfaces are investigated to measure the fineness of the surface. The more the number of faces, the finer is the reconstructed surface. The data in Table III shows the number of faces of the reconstructed surface for the corresponding input image for each of the algorithms. A graphical representation is given in Fig. 8. Among the selected algorithms, BPA reconstructs the minimum number of faces as it only creates a small collection of surfaces from point clouds. Alpha shape and PSR algorithms created more surfaces than BPA.

TABLE III. NUMBER OF 3D RECONSTRUCTED FACES OF EACH MODEL BASED ON SELECTED ALGORITHMS

Model	PSR	Alpha Shape	BPA
1	138666	226709	36429
2	224968	208791	51778
3	240180	280285	78132
4	136416	91976	7756

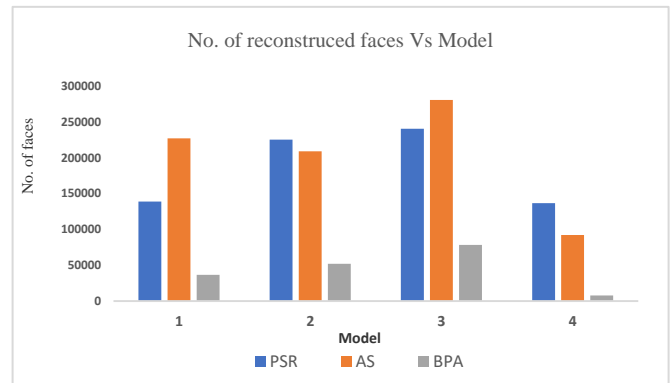


Fig. 8. Number of 3D Reconstructed Faces

#### V. CONCLUSION

A framework for the reconstruction of a 3D surface has been presented in the paper to reconstruct photorealistic 3D surfaces of irregularly formed objects from the set of their digital images to attain 3D point clouds that can be sparse, followed by dense clouds. The topology of the surface generated by these algorithms is based on the density of point clouds. Performance study of algorithms (PSA, Alpha shape, and BPA) has been carried out by observing their behavior with varying values of control parameters (octree depth, alpha value, and ball radius), on the amount of time taken and obtained quality of the reconstruction. Alpha shape and PSR algorithms performed better in creating more faces and hence a finer reconstructed 3D surface. To build a fine, feature-preserved, and accurate surface from the point cloud, several issues and constraints need to be considered. Noisy, sparse, and non-uniform points make the surface reconstruction process challenging. More research reconstructing finer surfaces using image features will be done in the future.

#### ACKNOWLEDGMENT

This research did not receive a specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

#### REFERENCES

- [1] S. P. Lim and H. Haron, "Surface reconstruction techniques: A review," *Artif. Intell. Rev.*, vol. 42, no. 1, pp. 59–78, 2014.
- [2] A. Alqudah, "Survey of Surface Reconstruction Algorithms," *J. Signal Inf. Process.*, vol. 05, no. 03, pp. 63–79, 2014.
- [3] H. Edelsbrunner and E. P. Mucke, "Three-dimensional Alpha Shapes 1 Introduction," *ACM Trans. Graph.*, vol. 13, no. 1, pp. 43–72, 1994.
- [4] F. Bernardini, J. Mittleman, H. Rushmeier, C. Silva, and G. Taubin, "The ball-pivoting algorithm for surface reconstruction," *IEEE Trans. Vis. Comput. Graph.*, vol. 5, no. 4, pp. 349–359, 1999.
- [5] N. Amenta, "The Crust Algorithm for 3D Surface Reconstruction from Scattered Points," *Proc. Conf. Comput. Geom. (SCG'99)*, pp. 423–424, 1999.
- [6] N. Amenta, S. Choi, and R. K. Kolluri, "The power crust," *Proc. Symp. Solid Model. Appl.*, pp. 249–260, 2001.
- [7] N. Amenta, S. Choi, T. K. Dey, and N. Leekha, "A simple algorithm for homeomorphic surface reconstruction." *Internat. J. Comput. Geom. & Appl.*, 12:125–141, 2002.
- [8] T. K. Dey and S. Goswami, "Tight cocone: a water-tight surface reconstructor." *J. Comput. Inf. Sci. Eng.* 3, no. 4, pp. 302–307, 2003.
- [9] T. K. Dey and S. Goswami, "Provable surface reconstruction from noisy samples." In *Proceedings of the twentieth annual symposium on Computational Geometry*, pp. 330–339. 2004.
- [10] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson Surface Reconstruction," in *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, Cagliari, Sardinia, Italy, vol. 256, pp. 61–70, 2006.
- [11] M. Kazhdan and H. Hoppe, "Screened poisson surface reconstruction," *ACM Trans. Graph.*, vol. 32, no. 3, pp. 1–13, 2013.
- [12] Khatamian and H. R. Arabnia, "Survey on 3D surface reconstruction," *J. Inf. Process. Syst.*, vol. 12, no. 3, pp. 338–357, 2016.
- [13] R. Tang, S. Halim, and M. Zulkepli, "Surface Reconstruction Algorithms: Review and Comparison," no. 1984, p. 22, 2013.
- [14] V. Chavan, Jimit Shah, M. Vora, M. Vora, and S. Verma, "A Review on Feature Extraction Methods for Image Analysis," *SSRN Electron. J.*, pp. 7–10, 2020.
- [15] Banerjee and D. Mistry, "Comparison of Feature Detection and Matching Approaches: SIFT and SURF," *GRD Journals-Global Res. Dev. J. Eng.*, vol. 2, pp. 7–13, 2017.
- [16] R. Sharma and P. Abrol, "Impact of Distortions on the Performance of Feature Extraction and Matching Techniques," *Lect. Notes Electr. Eng.*, vol. 701, pp. 357–367, 2021.
- [17] B. B. Swapnali and K. S. Vijay, "Feature Extraction Using Surf Algorithm for Object Recognition," *Int. J. Tech. Res. Appl.*, vol. 2, no. 4, pp. 197–199, 2014.
- [18] Maiti and D. Chakravarty, "Performance analysis of different surface reconstruction algorithms for 3D reconstruction of outdoor objects from their digital images," *Springerplus*, vol. 5, no. 1, 2016.
- [19] J. E. Nunes Masson and M. R. Petry, "Comparison of Algorithms for 3D Reconstruction," *19th IEEE Int. Conf. Auton. Robot Syst. Compet. ICARSC 2019*, pp. 1–6, 2019.
- [20] Z. Huang, Y. Wen, Z. Wang, J. Ren, and K. Jia, "Surface Reconstruction from Point Clouds: A Survey and a Benchmark," pp. 1–27, 2022.

# Classifying Weather Images using Deep Neural Networks for Large Scale Datasets

Shweta Mittal, Om Prakash Sangwan

Computer Science and Engineering, Guru Jambheshwar University of Computer Science and Engineering, Hisar, India

**Abstract**—Classifying weather from outdoor images helps prevent road accidents, schedule outdoor activities, and improve the reliability of vehicle assistant driving and outdoor video surveillance systems. Weather classification has applications in various fields such as agriculture, aquaculture, transportation, tourism, etc. Earlier, expensive sensors and huge manpower were used for weather classification making it very tedious and time-consuming. Automating the task of classifying weather conditions from images will save a huge time and resources. In this paper, a framework based on the transfer learning technique has been proposed for classifying the weather images with the features learned from pre-trained deep CNN models in much lesser time. Further, the size of the training data affects the efficiency of the model. The larger amount of high-quality data often leads to more accurate results. Hence, we have implemented the proposed framework using the spark platform making it scalable for big datasets. Extensive experiments have been performed on weather image dataset and the results proved that the proposed framework is reliable. From the results, it can be concluded that weather classification with the InceptionV3 model and Logistic Regression classifier yields the best results with a maximum accuracy of 97.77%.

**Keywords**—Weather classification; big data; transfer learning; deep learning; Sparkll; convolutional networks

## I. INTRODUCTION

Weather classification is a task of classifying weather conditions by looking at the weather data. Accurate weather classification enables users to organize day to day activities for instance deciding what type of clothes to wear, planning outdoor travel or sports activity, solar technologies, etc. [1]. It plays a critical role in various other fields such as in agriculture, it helps farmers to decide which fertilizers or pesticides to use, whether to turn off the sprinklers etc. based on the outdoor weather conditions. Accurate weather classification also helps in improving reliability of vehicle assistant driving and outdoor video surveillance systems [2]. Timely prediction of severe weather conditions helps saving people life and property. It helps to avoid road accidents, train derailment and ship collisions caused due to severe weather conditions like rain, fog, storm and snow [3]. It also assists in predicting natural calamities, thereby helping in saving people's life to a greater extent. Thus, timely and accurate classification of weather images is of utmost important to society and needs to be explored further by the researchers and meteorologists.

Convolutional Neural Network (CNN) is the commonly used deep learning architecture now-a-days to classify image

datasets. It comprises multiple convolutional and pooling layers for better hierarchical representation and a Fully Connected Layer at the end for classification. Numerous variants of CNN such as ResNet, Inception V3, Exception, VGG-16, and VGG-19, etc. have been developed by researchers in past for improving the accuracy of networks. Several features such as sky, shadow, reflection, haze, contrast, etc. are present in images which are used by the CNN for weather classification.

In this paper, a framework based on transfer learning techniques has been implemented for classifying the weather images into its appropriate category with the help of features extracted from pre-trained deep CNN models. Automating the classification will not only save time and resources but also aids in increasing the reliability of the process. Section II highlights the work done by researchers in the field of weather image classification. Limitation and challenges of existing technologies has been presented in Section III. Section IV defines the proposed framework. In Section V, the hardware and software required to implement the proposed framework have been discussed. The experimental results have been described in Section VI and the conclusion and future work has been presented in Section VII of the paper.

## II. RELATED WORK

In the past few years, a significant amount of work has been done by researchers to resolve the issue of weather classification. As the weather dataset is quite huge, researchers have implemented big data technologies such as Hadoop and Spark to process weather datasets. Alam and Ajmad proposed architecture for parallel and distributed processing of big Weather datasets using Hadoop and Map Reduce platforms [4]. Yang et al. used big data technology to implement a regression model for predicting weather conditions with a large amount of weather data [5]. Ismail et al. build a big data framework for predicting weather temperature using Map Reduce technique and the proposed framework had the advantage of improved efficiency and scalability [6].

Dhoot et al. implemented the ARIMA technique and Kalman filter to forecast the weather using a spark framework and it was found that the ARIMA model with spark resulted in a significant reduction of execution time [7]. Spark was also used by Jayanthi and Sumathi to analyze weather datasets using ipython notebook and the proposed framework overcome the drawback of Hadoop in terms of processing speed [8]. Sudmant et al. studied the values and limitations of big data for fair weather prediction and emphasis on the need for technological advances for sustainable development [9].

In past, numerous authors have experimented with different architectures of neural networks to forecast weather conditions and the results were quite promising. Reddy and Babu surveyed various models for weather forecasting and it was found that Feed Forward Neural Networks (FFNN) and Time Delay Recurrent Neural Networks (TDNN) are well suited for short-term and long-term forecasting respectively [10]. Ren et al. too surveyed deep learning-based weather forecasting models and compared their performance with conventional NWP techniques [11].

Tran et al. reviewed neural networks for air temperature forecasting and found that Multi-Layer Perceptron (MLP), FFNN, Ward Style ANN, Radial Basis Function (RBF), Generalized Regression Neural Network (GRNN), etc. were widely used networks whereas few studies also used deep learning networks which provide encouraging results [12]. A comparison has been done between traditional and emerging techniques used in weather forecasting and it was concluded that for classification, CART, Adaboost and XGBoost gave the best results while for prediction, Linear Regression gave better performance [13].

Sharma and Ismail implemented 19 layers deep CNN using Keras and Tensorflow library for classifying the weather images into 4 classes (rain, shine, sunshine, and cloud), and the model provided 94% accuracy [14]. Wang and Li applied the fusion of DenseNet and ResNet CNN for classifying the weather images into nine classes and it was concluded that the output of the proposed integrated model was better than implementing the basic models alone [15]. Xiao et al. build a weather dataset comprising 6877 images of 11 different categories and proposed MeteCNN, a deep CNN model for classification [16]. The proposed model gave 92% accuracy and its performance was found to be superior to VGG19, VGG16, ResNet18, ResNet34, and MobileNet models.

Elhoseiny et al. studied the CNN layer's performance for two pre-trained models i.e. ImageNet-CNN and Weather-Trained CNN to classify weather images and the proposed work outperformed state-of-art methods [2]. Ibrahim et al. proposed a WeatherNet framework that combines both weather and visual conditions and comprises 4 parallel deep CNN models a) NightNet to classify dawn, night, day, and dusk b) GlareNet to classify Glare and No Glare c) PrecipitationNet to detect clear, rain and snow d) FogNet to detect fog and no fog [17]. Xia et al. introduced the ResNet-15 model to classify the "Weather-Dataset 4" dataset consisting of weather images into four classes and it was concluded that the proposed model performed superior to the traditional ResNet50 model [18].

Dhananjaya et al. discussed the application of weather classification for autonomous driving and build a dataset for the classification of weather, light level, and street type into nine classes [19]. An active learning framework has been implemented to reduce redundancy in the dataset and the ResNet50 network has been used for classification. Guerra et al. too build a weather dataset comprising 5500 weather images of five different categories and further implemented CNN for feature extraction using Matlab and Caffe library. The experimental results proved that all the variants of ResNet performed better than VGG and CaffeNet [3]. Matteussi et al.

implemented a Deep Learning algorithm on the BigDL framework using spark and Tensorflow library and the results concluded that ResNet performed better than VGG in terms of speed while LeNet achieved higher accuracy [20].

Zhao et al. too implemented the CNN-RNN approach for multi-label weather recognition to extract co-related visual features [21]. CNN model is extended with the attention model for feature extraction and the RNN model is used to model the dependency among labels. The experimental result proved that the proposed model is better than VGGNet. Li et al. combined CNN features with other features such as haze, contrast, brightness, etc. to construct a high dimensional vector for multi-class weather classification and the proposed model is proved to be effective and inspiring [22].

Zhang and Ma proposed Multiple Kernel Learning (MKL) for multi-class weather image classification and the results proved that the proposed technique gave better results as compared to SVM and AdaBoost [23]. Kang et al. presented a framework for weather image recognition to classify weather images into three categories: hazy, snowy, and rainy. The performance of GoogleNet, Alexnet, and the Multiple Kernel-based Learning approach was compared and it was found that GoogleNet slightly outperformed Alexnet [24].

Chen et al. proposed a weather recognition system using MKL and active learning techniques for automatically labeling the weather images into sunny, cloudy, and overcast categories, and the proposed system achieved better performance [25]. Lu et al. proposed an approach using a collaborative learning strategy to classify a dataset of 10k images into two categories: sunny and cloudy [26]. The proposed approach has been compared with SVM, Adaboost, LLC, and ScSPM techniques and the results proved the efficiency of the proposed approach.

Chu et al. build a large-scale dataset to estimate weather from images and demonstrated the relationship between different visual and weather features [27]. Random Forest has been implemented to classify the images and the results were found to be satisfactory for sunny, cloudy, and snowy classes. For rainy and foggy, the model did not give good results. Li et al. constructed a decision tree-based SVM classifier for extracting features to automatically acquire weather conditions from images using C++ and OpenCV [28].

Wang et al. introduced a new method for multi-class weather classification which builds feature vectors from the combination of real-time weather factors and dark channel feature value derived from images [29]. The proposed model is implemented using the Adaboost classifier on 5k images and the results demonstrated that the model performed better. Transfer learning has been applied by Challa and Vaishnav using the pre-trained ResNet50 and DenseNet161 models to classify the weather images [30].

As per the review of the work done by researchers, it can be concluded that Multiple Kernel Learning and pre-trained Neural Networks such as AlexNet, ResNet, VGG, and GoogleNet have been widely used by researchers for the classification of weather images. Big Data frameworks such as Spark and Hadoop have also been explored to process big datasets.

### III. CHALLENGED AND LIMITATIONS OF EXISTING APPROACHES

Many existing weather classification techniques make use of expensive sensors and huge manpower for classifying weather images. These methods rely on human observations, thus are more prone to errors and are very time-consuming. Numerical Weather Prediction (NWP) models are also being widely used for forecasting the weather conditions, but the techniques are quite expensive and rely on the power of supercomputers for data processing.

Recent work on weather classification includes classification of weather conditions from images. CNN has an advantage of automatically detecting the important features from images without any human intervention. For getting the optimized results, CNN models require huge amount of processing power and large datasets for training. Further, training weather images using CNN models also require tuning number of hyper-parameters such as number of convolutional and max-pooling layers, kernel size, regularization techniques, etc.

To overcome the above mentioned issues, transfer learning techniques are applied to transfer the knowledge learned from pre-trained deep CNN models to our weather image dataset. As transfer learning utilizes the features extracted from pre-trained CNN models, training time is considerably reduced. It too eliminates the need for tuning of number of hyper-parameters. Transfer learning has proved its potential in number of applications such as natural language processing, sentiment classification, text classification, spam email detection, video classification, drug efficacy classification, etc. [31].

Further, the size of weather images generated is also huge. Existing weather classification techniques exploiting the transfer learning techniques do not make use of big data technologies as well as lacks scalability. Thus, there is a need of weather classification framework which can scale on number of machines and can support big weather image datasets.

### IV. PROPOSED METHODOLOGY

To overcome the limitations discussed in previous section, a weather image classification framework has been proposed and implemented using spark platform. Fig. 1 describes the proposed framework which comprises of three steps: image loading, feature learning and multi-class classification. Firstly, weather images are loaded into the spark via the readSchema() method of the ImageSchema library and a dataframe is created. Afterward, image data is given as input to the pre-trained deep CNN model to automatically learn hierarchical feature representation. At this stage, either we can freeze all the layers of the pre-trained models or some of its layers are unfreeze and fine-tuned to learn the desired features [32]. The final step is to use the learned features from the deep CNN models as input to the classifier model which classifies the outdoor weather image into its appropriate category.

In step 2, 5 deep CNN networks i.e. InceptionV3, Xception, ResNet50, VGG16, and VGG19 have been explored for feature learning and experiments have been conducted to determine the model which gave the highest accuracy. VGG16 and

VGG19 is 16 and 19 layered deep neural network whereas ResNet50 is 50 layered deep Convolutional neural networks. Inception V3 is also a very powerful neural network which is 42 layers deep and has an advantage of performing multiple different transformations. Xception is a 71 layers deep neural network and incorporates depth-wise separable convolutions. All the layers from the pre-trained CNN models have been frozen in this case study.

In the last stage of proposed framework, one of the two classifiers i.e. Logistic Regression and Random Forests has been implemented for classification of weather images. Logistic Regression is a supervised ML technique which uses a sigmoid activation function for solving classification problems. To reduce the over-fitting and generalization error, the Logistic Regression classifier is trained with elastic net regularization. Two hyper-parameters for Logistic Regression need to be optimized; alpha ( $\alpha$ ) which is used to assign the weight to both L1 and L2 penalty and lambda ( $\lambda$ ), a regularization parameter that defines the balance between minimizing the training error and model complexity. If  $\alpha$  is 0, the model follows L2 regularization while if  $\alpha$  is 1, the model follows L1 regularization.

Random Forests is another popular supervised machine learning technique based on ensemble learning which combines the results of a large number of weak decision trees to form a robust classifier. As shown in Fig. 2, 10 different combinations of the networks are possible and the same has been implemented using the Sparkdl library. Extensive experiments have been performed to determine the CNN model and classifier which provided the maximum accuracy for weather image classification.

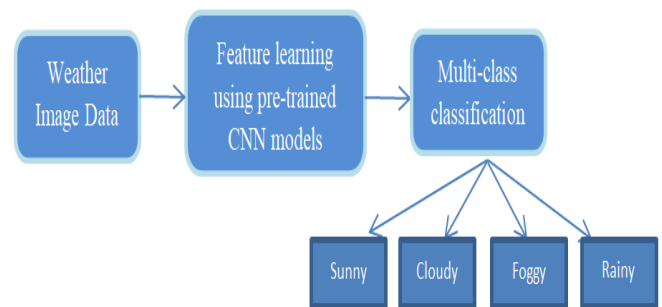


Fig. 1. Proposed framework for weather image classification.

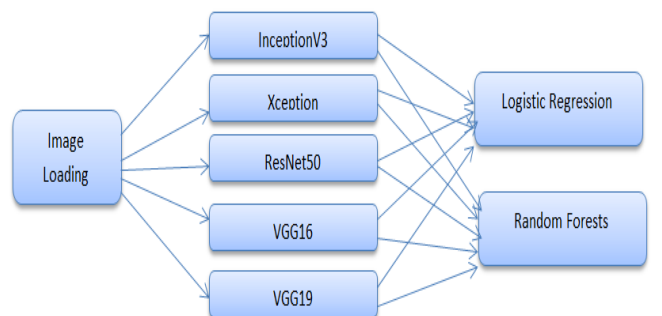


Fig. 2. Proposed architecture for weather classification.



## V. EXPERIMENTAL SETUP

To evaluate the efficiency of the proposed framework, Multi class weather image classification datasets freely available on Kaggle.com has been used [33]. Dataset comprises 1125 images of four different categories i.e. cloudy, rainy, shiny, and sunrise. Sample images from each target class have been shown in Fig. 3. Categorization of target classes for the dataset has been presented in Fig. 4.

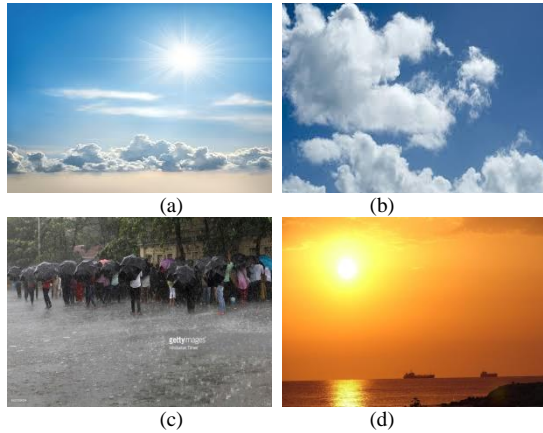


Fig. 3. Outdoor weather images (a) shine (b) cloudy (c) rainy (d) sunrise.

For the experimental setup, spark version 2.3.0 has been set up on a local PC (i.e. driver and the executor are set up on the same local machine) with 16GB RAM (12 GB RAM is allocated to spark) and 1 TB hard disk. Python version 3.7 and Java version 1.8 has been installed as shown in Table I. Further, there are more than 150 parameters in Spark which needs to be tuned for any spark application to achieve optimized results. There is no default value of a parameter that fits every application; values of these parameters vary with workload and the type of application.

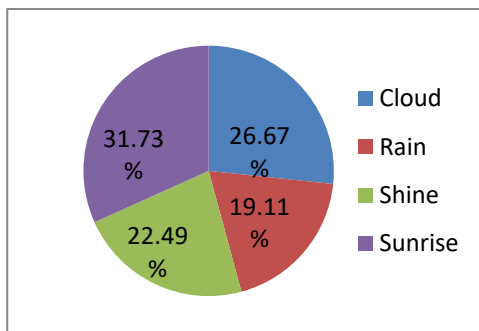


Fig. 4. Categorization of target class.

TABLE I. LOCAL PC CONFIGURATION

Parameter	Value
RAM	16GB
Hard Disk	1TB
OS	64-bit Windows 10
Spark version	2.3.0
Java	1.8
Python	3.7

Optimal tuning of spark parameters i.e. driver memory, driver cores, executor cores, executor memory, number of partitions etc. improves the performance of spark application to the greater extent [34,35,36,37]. As size of data is not very large, 4g memory is sufficient to store the data as well as to run the CNN models with Logistic Regression as classifier. While for Random Forest classifier, minimum 8g driver memory is required. Number of cores allocated to driver/executor too impacts the application's performance. If a cluster has n number of cores, number of partitions for input data should not be less than n as it might lead to under-utilization of resources. Few cores will be left idle and resources would not be utilized efficiently. Thus, number of input partition should either be equal or greater than number of cores. In this research work, number of cores and partitions has been set to 2 and 4, respectively.

To evaluate the efficiency of proposed framework, following metrics have been used:

- 1) Accuracy: It can be classified as the ratio of truly predicted samples to the total number of predictions.
- 2) Precision: It can be defined as the ratio of truly predicted positive values to the total number of positive predictions.
- 3) Recall: It can be defined as the ratio of true positives and total positive samples.
- 4) F1 score: Calculated as the harmonic mean of precision and recall.
- 5) Training Time: Time required for training the model.

## VI. RESULTS AND DISCUSSION

For Logistic Regression classifier, impact of regularization parameter, alpha and lambda on model's accuracy has been studied. Accuracy values for different value of alpha have been observed and results are shown in Table II. It can be interpreted that accuracy of the Logistic Regression model trained using ResNet50 model is maximum when the value of alpha lies between 0 and 0.5.

TABLE II. IMPACT OF ALPHA ON MODEL'S ACCURACY

Alpha	Accuracy
0	97.28
0.1	97.28
0.2	97.28
0.3	96.8
0.4	96.8
0.5	97.28
0.6	96.8
0.7	96.38
0.8	95.92
0.9	95.92
1	96.28

Further, 5 fold Cross-Validation method has been used to determine the optimal lambda value of Logistic Regression.

Numerous values of  $\lambda$  i.e. (0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 1.0) have been examined and the results concluded that the model achieves higher accuracy for  $\lambda$  value of 0.01. Thus, for implementing the Logistic Regression classifier,  $\alpha$  and  $\lambda$  has been set to 0.5 and 0.01, respectively.

For Random Forest classifier, the number of trees in forests impact performance of classifier to a greater extent. Thus, for different values of trees, its impact on the accuracy and run time of an algorithm has been studied and results are shown in Table III. It can be interpreted that till a certain value i.e. 150, the accuracy value of the model increases with an increase in the number of trees in the forests. Afterwards, the accuracy value decreases. Random Forest classifier provided the best accuracy value of 97.73% with 150 trees in the forests. It can also be observed that the increase in the number of trees has no significant impact on the training time of an algorithm.

Train and test data have been divided in the ratio 80:20. Table IV describes the accuracy, F1 score, precision, recall, and training time for 10 different possible architectures specified in Fig. 2. It can be interpreted that with the InceptionV3 pre-trained model and Logistic Regression classifier, the accuracy value is maximum i.e. 97.77%. Factorized convolutions, grid size reduction, asymmetric convolutions and use of auxiliary classifiers are some the significant features of InceptionV3 model resulting in its higher accuracy value. Instead of getting deeper, InceptionV3 extracts features in more depth.

TABLE III. IMPACT OF NUMBER OF TREES ON THE PERFORMANCE OF PROPOSED NETWORK

No. of trees	Accuracy	F1 score	Precision	Recall	Time
10	94.57	94.55	94.68	94.57	14 min 24 sec
50	96.39	96.38	96.4	96.38	16 min 3 sec
100	97.28	97.29	97.41	97.28	14 min 35 sec
125	97.29	97.29	97.32	97.28	13 min 58 sec
150	97.73	97.73	97.82	97.73	13 min 1 sec
200	96.38	96.4	96.49	96.38	14 min 5 sec

TABLE IV. OUTPUT OF PROPOSED FRAMEWORK FOR NUMEROUS PRE-TRAINED DEEP CNN MODELS

	Accuracy	F1 score	Precision	Recall	Time
ResNet50+LR	97.28	97.29	97.33	97.28	7 min 35 sec
VGG16+LR	94.57	94.61	94.83	95.57	10 min 12 sec
VGG19+LR	94.12	94.13	94.31	94.12	6 min 35 sec
InceptionV3+LR	97.77	97.74	97.75	97.74	6 min 42 sec
Xception+LR	95.02	95	95.11	95.02	8 min 2 sec
ResNet50+RF	97.73	97.73	97.82	97.73	13 min 1 sec
VGG16+RF	96.38	96.35	96.46	96.38	28 min 1 sec
VGG19+RF	93.66	93.64	93.62	93.66	26 min 45 sec
InceptionV3+RF	91.4	91.5	91.7	91.4	24 min 14 sec
Xception+RF	90.04	89.98	90.32	90.04	50 min 50 sec

Further, the second highest accuracy value i.e. 97.73% has been achieved with the ResNet50 pre-trained model and Random Forests classifier. ResNet50 pre-trained model with the Logistic Regression classifier too gave promising results. The Xception model trained using the Random Forests classifier provided the least accurate value of 90.04%. It can also be concluded that the time taken by the framework for training the dataset with the Logistic Regression classifier is quite low compared to the Random Forests.

Table V demonstrates the confusion matrix for weather image classification for dataset 1 with Inception V3 pre-trained CNN model and Logistic Regression classifier, where the row represents the actual labels and the column represents the predicted labels. As shown in Table III, the total number of test images is 221. The number of test images for classes cloudy, foggy, rain, shine, and sunshine is 45, 64, 47, and 65, respectively. Out of 221, 216 images are predicted true by the classifier.

TABLE V. CONFUSION MATRIX FOR INCEPTION V3 LOGISTIC REGRESSION MODEL

	Cloudy	Rainy	Shine	Sunrise	Total
Cloudy	44	0	1	0	45
Rainy	1	63	0	0	64
Shine	0	0	45	2	47
Sunrise	0	0	1	64	65
Total	45	63	47	66	221

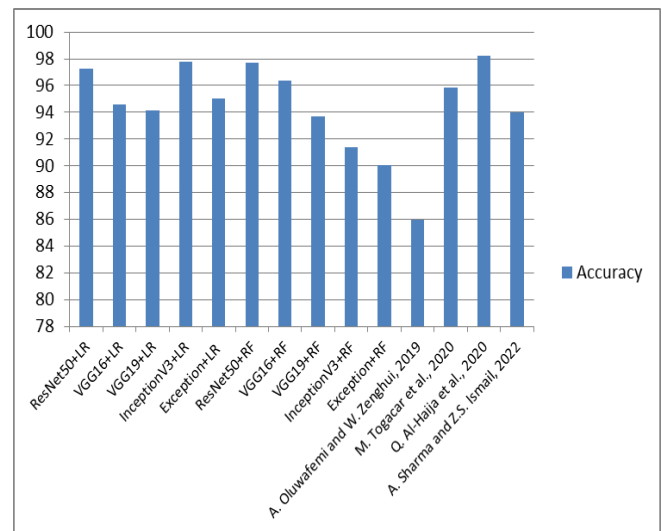


Fig. 5. Comparison of the proposed framework with the existing techniques.

The performance of proposed framework has been compared with the existing techniques utilizing the same weather image dataset for classification. Fig. 5 shows the comparison of accuracy value of the proposed framework with the four existing approaches. Our proposed framework yields the maximum accuracy value of 97.77% when Inception V3 model and Logistic Regression classifier (InceptionV3+LR) has been used. Models proposed by Oluwafemi and Zenghui (2019), Togacar et al. (2020), and Sharma and Ismail (2022) had lower accuracy values of 86%, 95.85% and 94%

respectively [14,38,39]. Model proposed by Al-Haija et al. (2020) had the accuracy value of 98.22%, thus performed slightly better than the proposed InceptionV3+LR model in terms of accuracy [40]. Further, Model proposed by Al-Haija et al. had precision and sensitivity values of 96.5% and 96.4%, respectively while the proposed InceptionV3+LR model has better precision and sensitivity values i.e. 97.75% and 97.74% respectively. Hence, InceptionV3+LR model has better ability to capture false negatives and true positives than existing techniques. Time taken by the proposed framework for training is also very less i.e. 6 min 42 sec.

## VII. CONCLUSION

Weather Prediction is a very vital activity performed before numerous day-to-day activities. Deep CNN is a deep learning technique widely used to classify image datasets but the techniques require very huge datasets and high computing devices for achieving optimal results. Transfer learning is proven to be effective in scenarios with smaller datasets and the unavailability of high computing devices. It helped in minimizing the learning time by utilizing the features learned from the previously trained deep CNN models.

In this paper, a framework based on transfer learning has been proposed for classifying the weather conditions from outdoor images and the performance of various CNN models has been compared for accuracy, precision, recall, f1 score, and training time. From the results, it can be observed that with the InceptionV3 pre-trained model and Logistic Regression classifier, the accuracy value is maximum i.e. 97.77%. Further, it can be concluded that the time taken by the framework for training the dataset with the Logistic Regression classifier is quite low compared to the Random Forests. The comparison has also been done with the existing techniques and the results proved the efficiency of the proposed framework. Further, the proposed framework is scalable and has support for big datasets. In the future, some layers from pre-trained CNN models can be un-freeze and fine-tuned to learn better hierarchical representation.

## REFERENCES

- [1] B. Zhao, X. Li, X. Lu, and Z. Wang, "A CNN-RNN architecture for multi-label weather recognition," *Computer Science, Computer Vision and Pattern Recognition*, 2019.
- [2] M. Elhoseiny, S. Huang, and A. Elgammal, "Weather classification with deep convolutional neural networks," *International Conference on Image Processing (ICIP)*, IEEE, Canada, 2015.
- [3] J. C. Villarreal Guerra, Z. Khanam, S. Ehsan, R. Stolkin, and K. McDonald-Maier, "Weather classification: A new multi-class dataset, data augmentation approach and comprehensive evaluations of convolutional neural networks," *NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*, pp. 305-310, 2018.
- [4] M. Alam and M. Ajmad, "Weather forecasting using parallel and distributed analytics approaches on big data clouds," *Journal of Statistics and Management Systems*, vol. 12 no. 4, pp. 791-799, 2020.
- [5] N. Yang, L. Westfall, and P. Dalvi, "A weather prediction model with big data," *Proceedings of Student-Faculty Research Day, CSIS*, 2018.
- [6] K. A. Ismail, M.A. Majid, J. Mohamed Zain, and N.A. Abu Bakar, "Big data prediction framework for weather temperature based on MapReduce algorithm," *IEEE Conference on Open Systems*, 2016.
- [7] R. Dhoot, S. Agrawal, S. Kumar, and M. Sivagami, "Implementation and analysis of Arima model and kalman filter for weather forecasting in spark computing environment," *3rd International Conference on Computing and Communications Technologies (ICCT)*, IEEE, 2019.
- [8] D. Jayanthi and G. Sumathi, "Weather data analysis using spark – An In-memory computing framework," *International Conference on Innovations in Power and Advanced Computing Technologies [i-PACT2017]*, IEEE, 2017.
- [9] A. Sudmant, V. Viguie, Q. Lepetit, L. Oates, A. Datey, A. Gouldson, and D. Watling, "Fair weather forecasting? The shortcomings of big data for sustainable development, a case study from Hubballi-Dharwad, India," *Sustainable Development*, vol. 24, no. 6, pp. 1237-1248, 2021.
- [10] P.C. Reddy and A.S. Babu, "Survey on weather prediction using big data analytics," *Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, IEEE, 2017.
- [11] X. Ren, X. Li, K. Ren, J. Song, Z. Xu, K. Deng, and X. Wang, "Deep learning-based weather prediction: A survey," *Big Data Research*, Elsevier, vol. 13, 2020.
- [12] T.T.K. Tran, S.M. Bateni, S.J. Ki, and H. Vosoughifar, "A review of neural networks for air temperature forecasting," *Water*, vol. 13, no. 9, 2021.
- [13] I. Gad and D. Hosahalli, "A comparative study of prediction and classification models on NCDC weather data," *International Journal of Computers and Applications*, vol. 44 no. 5, pp. 414-425, 2022.
- [14] A. Sharma and Z.S. Ismail, "Weather classification model performance: Using CNN, Keras-TensorFlow," *1st International Conference on Applied Computing & Smart Cities (ICACS21)*, vol. 52, 2022.
- [15] Y. Wang and Y. Li, "Research on multi-class weather classification algorithm based on multi-model fusion," *IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pp. 2251-2255, 2020.
- [16] H. Xiao, F. Zhang, Z. Shen, K., Wu, and J. Zhang, "Classification of weather phenomenon from images by using deep convolutional neural network," *Earth and Space Science*, vol. 8, 2021.
- [17] M.R. Ibrahim, J. Haworth, and T. Cheng, "WeatherNet: Recognizing weather and visual conditions from street-level images using deep residual learning," *Computer Science, Computer Vision and Pattern Recognition*, 2019.
- [18] J. Xia, D. Xuan, L. Tan, and L. Xing, "ResNet15: Weather recognition on traffic road with deep Convolutional neural network," *Advances in Meteorology, Hindwai*, 2020.
- [19] M.M. Dhananjaya, V.R. Kumar, and S. Yogamani, "Weather and light level classification for autonomous driving: Dataset, baseline and active learning," *Computer Science, Computer Vision and Pattern Recognition*, 2019.
- [20] K.J. Matteussi, B.F. Zanchetta, G. Bertoncello, J. Santos, J. Anjos, and C. Geyer, "Analysis and performance evaluation of deep learning on big data," *IEEE Symposium on Computers and Communications, Spain*, 2020.
- [21] X. Zhao, S. Qi, B. Zhang, H. Ma, W. Qian, Y. Yao, and J. Sun, "Deep CNN models for pulmonary nodule classification: model modification, model integration, and transfer learning," *Journal of X-Ray Science and Technology*, vol. 27, no. 4, pp. 615-629, 2019.
- [22] Z. Li, Y. Li, J. Zhong, and Y. Chen, "Multi-class weather classification based on multi-feature weighted fusion method," *2nd International Conference on Oil & Gas Engineering and Geological Sciences, IOP Conf. Series: Earth and Environmental Science*, 2020.
- [23] Z. Zhang and H. Ma, "Multi-class weather classification on single images," *International Conference on Image Processing (ICIP)*, IEEE, pp. 4396-4400, 2015.
- [24] L. Kang, K. Chou, and R. Fu, "Deep learning based weather image recognition," *International Symposium -on Computer, Consumer and Control (IS3C)*, pp. 384-387, 2018.
- [25] Z. Chen, F. Yang, A. Lindner, G. Barrenetxea, and M. Vetterli, "How is the weather: automatic inference from images," *19th IEEE International Conference on Image Processing*, pp. 1853-1856, 2012.
- [26] C. Lu, D. Lin, J. Jia, and C.K. Tang, "Two-class weather classification," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3718-3725, 2014.

- [27] W.T. Chu, X.Y. Zheng, and D.S. Ding, "Image2Weather: A large-scale image dataset for weather property estimation," Second International Conference on Multimedia Big Data (BigMM), IEEE, pp. 137-144, 2016.
- [28] Q. Li, Y. Kong, and S. Xia, "A method of weather recognition based on outdoor images," International Conference on Computer Vision Theory and Applications (VISAPP), pp. 510-516, 2014.
- [29] S. Wang, Y. Li, and W. Liu, "Multi-class weather classification fusing weather dataset and image features," In: Z. Xu, X. Gao, Q. Miao, Y. Zhang, J. Bu (eds) Big Data, Big Data, Communications in Computer and Information Science, vol. 945, Springer, Singapore, 2018.
- [30] S.V.D.M. Challa and H.K. Vaishnav, "Weather categorization using foreground subtraction and deep transfer learning," In: H. Saini, R. Sayal, R. Buyya, G. Aliseri (eds), Innovations in Computer Science and Engineering, Lecture Notes in Networks and Systems, vol. 103, Springer, Singapore, 2020.
- [31] K. Weiss, T.M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," Journal of Big Data, vol. 3, 2016.
- [32] R. Ribani and M. Marengoni, "A survey of transfer learning for Convolutional Neural Networks." 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), Brazil, pp. 47-57, 2019.
- [33] G. Ajayi, "Multi-class weather dataset for image classification," Mendeley Data, V1, 2018.
- [34] M. Rahman, J. Hossen, and C. Venkateshaiah, "SMBSP: A Self-Tuning Approach using Machine Learning to Improve Performance of Spark in Big Data Processing," 7th International Conference on Computer and Communication Engineering (ICCCE), IEEE, Malaysia, Sep. 2018.
- [35] P.N. Saipraveen and G.S. Nagaraja, "Parameter Tuning of Apache Spark based Applications for Performance Enhancement," International Research Journal of Engineering and Technology (IRJET), vol. 7, no. 5, May 2020.
- [36] Q. Chen, K. Wang, Z. Bian, I. Cremer, G. Xu, and Y. Guo, "Simulating Spark Cluster for Deployment Planning, Evaluation and Optimization," 6th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH), IEEE, Portugal, Jul. 2016.
- [37] H. Zhang, Z. Liu, and L. Wang, "Tuning Performance of Spark Programs," International Conference on Cloud Engineering (IC2E), IEEE, USA, Apr. 2018.
- [38] A.G. Oluwafemi and W. Zenghui, "Multi-Class Weather Classification from Still Image Using Said Ensemble Method," Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), IEEE, South Africa, 2019.
- [39] M. Togaçar, B. Ergen, and Z. Comert, "Detection of weather images by using spiking neural networks of deep learning models," Neural Computing and Applications, vol. 33, pp. 6147-6159, 2020.
- [40] Q.A. Al-Haija, A. Smadi, and S. Zein-Sabatto, "Multi-Class weather classification using ResNet-18 CNN for autonomous IoT and CPS applications," International Conference on Computational Science and Computational Intelligence (CSCI), IEEE, USA, 2020.

# Exploring College Academic Performance of K to12 IT and Non-IT-related Strands to Reduce Academic Deficiencies

Marilou S. Benoza<sup>1</sup>, Thelma Palaoag<sup>2</sup>

Infotech Development Systems Colleges, Inc, Ligao City, Philippines<sup>1</sup>  
University of the Cordilleras, Baguio City, Philippines<sup>2</sup>

**Abstract**—Improving students' academic performance is a significant concern among academics. Despite various strategies to improve academic performance, there are still a significant number of students who fail academically. This study sought to investigate the possible reasons for the academic deficiencies among Infotech Development Systems Colleges, Inc. Information Technology (IT) students in the IT and Non-IT-related strands, if their strand is significant to their academic performance in college, as well as formulate a solution based on the identified reasons and recommendations to reduce the negative academic remarks. The researchers employed survey questionnaires and interviews to conduct exploratory data analysis. Similarly, the Senior High School academic performance and actual grades of the respondents from AY 2018-2019 to First Semester of AY 2021-2022. On the tested hypothesis showed statistical significance ( $p < 0.05$ ) on IT-related strand. The study further reveals that the Non-IT-related strand has more students with academic deficiencies compared to IT-related strand and highlights a variety of reasons cited. Respondents cited misalignment of strand to current program, instructor not speaking clearly, unreliable internet connection, and failure to complete and submit an academic task as reasons for academic deficiencies. The researchers designed a model that can potentially eliminate academic inadequacies. The model takes into account both internal and external factors; for internal variable it includes effective time management, a positive attitude and mindset, prompt and punctual completion of requirements, and good study habits. While for the external factors, competent and student-friendly instructors, a stable, strong, and accessible internet connection, a conducive learning environment, relevant available resources and facilities, adaptation of limited face-to-face or hybrid classes, and alignment of SHS strand to college program of choice are recommended.

**Keywords**—Academic performance; exploratory research; reduce academic deficiencies; thematic analysis

## I. INTRODUCTION

Expanding our worldview through education broadens our perspective. Many nations, like the Philippines, are shaping education to be the most valuable resource a nation may possess. Philippine Basic Education introduced the K-12 Curriculum, which requires students to complete two years of senior high school before being eligible for college. Students who enroll in this program may select their own track or strand based on their preferences or areas of interest, or they may choose to pursue work for the remaining two years of

senior high school. Senior high school students follow a basic core curriculum and can specialize in one of four areas. These tracks are distinct fields of study, akin to college courses. The tracks are Academic, Technical-Vocational and Livelihood (TVL), Sports, and Arts and Design. Some of the most popular strand under Academic and TVL tracks are Science, Technology, Engineering, and Mathematics (STEM); Accountancy, Business, and Management (ABM); Humanities and Social Sciences (HUMSS); General Academic, Information and Communication Technology (ICT) and Home Economics (HE). Magdadaro [1] mentioned that having an ideal strand offer learners a sense of self-assurance and can help students to be passionate about their chosen career.

Academic performance is among the several components of academic success and key feature in education. [2], [3] Academic success and obtaining good grades are among the main goals in all levels of education while having positive outcomes both for the learners and educational systems. [4] The students' performance (academic achievement) plays an important role in producing the best quality graduates who will become great leaders and manpower for the country thus responsible for the country's economic and social development also one of the major factors considered by employers in hiring workers especially for the fresh graduates. Thus, students have to put the greatest effort in their study to obtain good grades and to prepare themselves for future opportunities in their career at the same time to fulfil the employer's demand. [5], [6] Many factors, including socioeconomic status, student temperament and motivation, peer, and parental support influence academic performance. [7] Factors affecting the academic performance of students are numerous and they can vary from nation to nation as well as even from person to person. As such, it would be really inadequate to investigate students' academic performance through a single-factor perspective [8] Students' performance also determines quality of education that will be passing to the students by the potential teachers at primary and secondary school levels. The good academic performance of students at the Senior High School is of paramount importance in every educational system. Meanwhile, numerous factors influence the academic performance of students and have been researched, but many problems persist. [9], [10].

There is a noticeable situation in which students enroll in college courses that are not associated with their track or

strand of senior high school. The mismatch rate is high between the strand of the learners during their senior high school and the course they enrolled in college. [11] Three mismatch characteristics affect students' academic outcomes: (1) a mismatch between expected and real grades, (2) a mismatch between expected and real levels of interest in studying, and (3) a mismatch between expected and real time for extracurricular activities at university. [12] The student's past performance in SHS has a significant effect on the student's self-efficacy in using information and communications technology (ICT)-related technologies in college. Computer self-efficacy also influences the students to achieve their academic goals in college; also, both past performance and computer self-efficacy have a strong influence on technology outcome expectations, however, the technology outcome expectations have a less significant effect on the students' achievement of their academic goals. [13] Low academic performance and adjustment to college are phenomena not new in the global education institutions. In fact, reported 60% of students who cannot adjust to college drop out early in school. The students who cannot establish good relationships with their friends, teachers and school administration, who do not like the school and the subjects have a higher tendency to be absent from school and to drop out of school. Also, that the specific causes of school dropouts include the difficulty of adjusting with the school curriculum. [14], [15], [16], [17] Moreover, mentioned that the problem of low academic achievement of students in the examinations is one of the most challenging problems that faces students as well as teachers. Low or weakness of the student's mark under the normal average in a study subject level as a result of a variety of reasons, including those related to the student himself, or those related to family, social and academic environment. Consequently, this may lead to frequent repetition of failure, despite their abilities that qualify them to get the best marks. [18].

In Infotech Development Systems Colleges, Inc. (IDSCI), it was discovered that a significant number of students in Information Technology programs have incomplete and failed remarks on their final grades, particularly in professional courses, as well as a significant number of dropouts, resulting in a decrease in program enrollment. In addition, students who took the non-IT-related strand in SHS are thought to have more incomplete and failed grades than those who took the IT-related strand.

The purpose of this study is to investigate the factors that contribute to the academic shortcomings of the respondents and to establish whether or not their strands have an impact on their academic performance. In addition to this, the researchers are interested in determining if the curriculum that they have chosen for their college is compatible with and advantageous to them. The researchers intend to carry out this study to serve as a guide to the institution, as well as for the students, particularly those who are enrolled and wishes to enroll in the Information Technology Department. The guidance will be presented in the form of a model with the goal of reducing the detrimental effect that it has on academic achievement. This model was designed using the causes that have been discovered for the academic deficiencies, as well as

the recommendations made by the students, which included the alignment of their SHS strand.

## II. METHODOLOGY

The exploratory research method was used in this study to determine ways to reduce academic deficiencies by examining respondents' college academic performance. This study was conducted among sixty-two (62) Information Technology (IT) students from Infotech Development System Colleges, Inc. and was divided into IT and Non-IT-related Strands based on their chosen strand. In order to gather enough information, the researcher used a survey questionnaire. The questionnaire was issued to eighty (80) target respondents, however only sixty (62) were retrieved. The proponent foregoes the available size with a confidence level of 90% and a margin of error of 5%.

After initiating the strand, the researchers identified the college academic achievement of the target respondents in order to determine the number of students that received INC and failed remarks. The researchers employed a questionnaire in which respondents were asked to assess their level of agreement with the criteria that had been devised to match the SHS Strand to the chosen college program. The questionnaire was distributed to respondents using Google Form, and the categories on the item on SHS strand matching on the college program, as well as if their strand was beneficial to their academic achievement in college, were listed to aid respondents in choosing the right decisions. In addition, weighted mean and standard deviation were utilized in order to identify and analyze the data.

The researcher also utilized the SHS and college academic performance to further determine if the respondent's strand has a significant relation to their academic performance in college. The research used the Pearson R statistical tool for the correlation. See (1). The SHS generated weighted average (GWA) was converted into a numerical value stipulated in the institution's Student Handbook. The numerical value was converted using a 6-point performance equivalent scale (6-Outstanding to 1-Failed).

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (1)$$

A separate survey questionnaire was also distributed only to respondents who had Incomplete or Failed remarks in order to determine the reason for their negative remarks. A 5-point Likert scale was used in both instruments. See Table I.

TABLE I. FIVE-POINT LIKERT SCALE FOR THE MATCHING AND BENEFITS OF SHS STRAND

Point Score	Range Interval	Descriptive Rating
		Level of Agreement
1	1.00-1.79	Strongly Disagree
2	1.80-2.59	Disagree
3	2.60-3.39	Neutral
4	3.40-4.19	Agree
5	4.20-5.00	Strongly Agree

Moreover, the researcher used thematic analysis to identify the recommendations by the respondents to reduce the academic deficiencies. Thematic analysis is the process of identifying patterns or themes within qualitative data. The goal of a thematic analysis is to identify themes, i.e. patterns in the data that are important or interesting, and use these themes to address the research or say something about an issue. This is much more than simply summarizing the data; a good thematic analysis interprets and makes sense of it [19].

A structured interview was conducted to collect the possible recommendations to reduce academic deficiencies. There were twenty-four (24) IT students who participated in the interview. These were the students who has failed or incomplete grades. Based on the respondents' recommendation, the analysis produced six (6) themes which included time management, positive attitude and mindset, instructor, requirement completion, study habits, and resumption of classes. The flowchart of the research process is presented on Fig. 1.

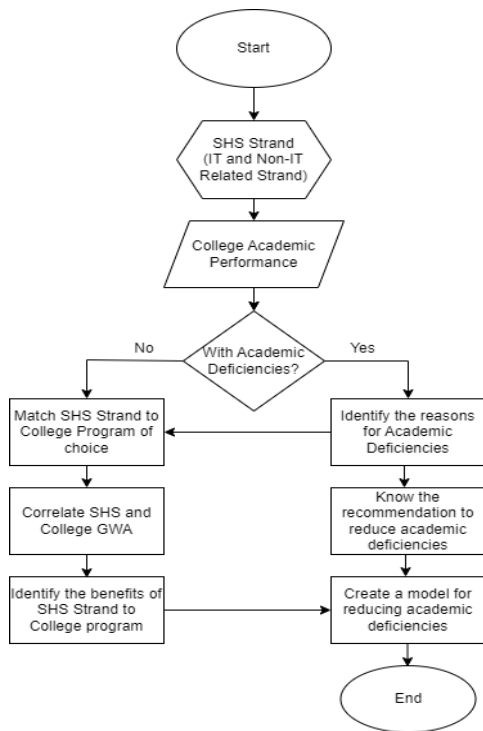


Fig. 1. Flowchart of the research process.

The researcher retrieved the instruments automatically through Google form. The academic performance of the respondents was measured using their actual grades from AY 2018-2019 to First Semester of the AY 2021-2022 taken from the Office of the Registrar with the permission of the IT students. The Student Profile Form was also used to identify the academic strand of the respondents. In this study, the respondents' Senior High School (SHS) strand were identified and categorized into IT Related Strand (ITRS) and Non-IT-Related Strand (NITRS). ITRS includes STEM and ICT while the NITRS includes ABM, HUMSS, GAS, Home Economics, Automotive, Agriculture, Arts and Design, Electrical, Dressmaking and Industry-al Arts. The meaning and

significance of the responses were interpreted through processes that included gathering, synthesizing, tabulating, and interpreting the meaning and significance of the responses. Tables and graphs were created to display the collected data. Data analysis was carried out using Microsoft Excel and Social Science Statistics Calculator. Mean, standard deviation, frequency, percentage, and correlation for descriptive analysis.

### III. RESULTS AND DISCUSSION

In Fig. 2 shows the academic remarks status of the respondents. The results show that 58% got passing remarks but despite the high rate of passing grade there were still 42% respondents who got incomplete or failed academic remarks. Fig. 2 also reveals that 19% belong to IT-related-strand while 23% also belong to non-IT-related strand students.

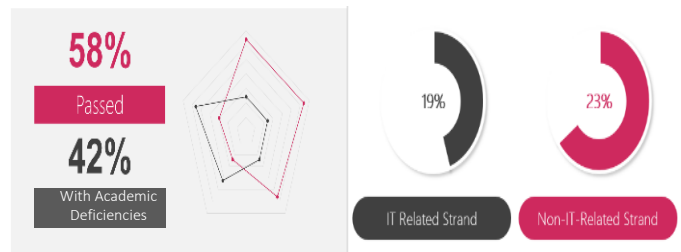


Fig. 2. Academic remarks status of the respondents.

Academic excellence is not universal among students; there are some students that despite of various strategies to improve academic performance, there are still number of students who fail academically. Failure in small doses is actually crucial in learning. However, when students completely fail academically, this means that they are unable to overcome the small failures over time to learn and grow and eventually succeed [20]. There are many causes of student failure, and these causes could be looked at in various ways. In this study, the researcher also would like to identify the reason of IT and non-IT-related strand for academic deficiencies in the area of subject matter, teacher factor, learning environment, motivation and attitude, and personal issues in order to provide sufficient solution based on the determined reasons for academic deficiencies. And also, to determine if their strand has an effect and beneficial to their college academic performance.

#### A. Reasons for Academic Deficiencies

1) *Subject matter*: Fig. 3 shows the reasons for academic deficiencies in terms of Subject Matter. In the course content was too difficult to understand, ITRS got a mean score of 3.33 with SD of 1.12 interpreted as neutral while NITRS got a mean score of 3.13 with SD of 1.13 interpreted as neutral. Too much course content/information was presented at a time making it difficult to be absorbed, ITRS yielded a mean score of 3.67 with SD of 0.71 interpreted as agree while NITRS gained a mean score of 3.47 with SD of 1.13 interpreted as agree. On the prerequisite competencies were not well developed in the preceding year level, ITRS obtained a mean score of 3.67 with SD of 1.00 interpreted as agree. On the other hand, NITRS obtained a mean score of 3.13 with SD of 1.25 and was

interpreted as neutral. In terms of the SHS strand undergone was not aligned with the present program being taken up, ITRS received a mean score of 2.78 with SD of 1.09 contrary to NITRS which received a mean score of 3.60 with SD of 1.40 interpreted as agree.

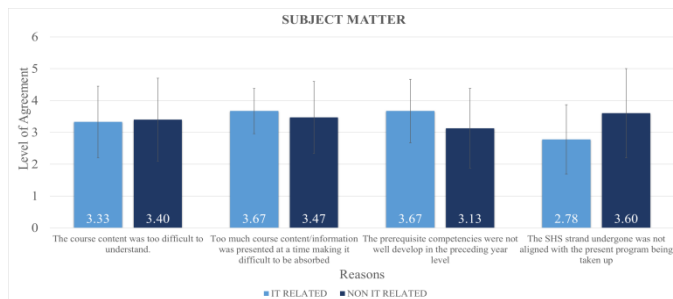


Fig. 3. Reasons for academic deficiencies in terms of subject matter.

According to the data, misalignment of the SHS strand to the enrolled program is one of the reasons why non-IT-related strand respondents received incomplete or failed remarks, as opposed to IT-related strand respondents. NITRS respondents mentioned that the topic presented in their current program is too different from the topics and discussion in their SHS strand, and they are finding it difficult to acclimatize. They also said that enrolling in an IT-related strand will make it easier for them to understand the lecture and other class activities. Moreover, the ITRS and NITRS both agreed that a fast-paced discussion was also one of the reasons for academic deficiencies. They have a hard time absorbing information that is delivered at the same time since it does not give them enough time to comprehend and learn the lesson. ITRS further stated that in the previous year's level, the required competences for the next course were not fully developed. Lack of competence causes a lack of confidence and willingness to complete tasks and activities. Students with strand mismatch lag behind others as it is more difficult for them to follow the unfamiliar subjects of their course. Students suffering from strand mismatch are also deprived of interpersonal relationships as forming them is directly related to academic engagement. It also has various indirect effects from these relationships to cognitive and behavioral participation [21].

2) *Teacher factor*: Fig. 4 shows the reasons for academic deficiencies in terms of the teacher factor. On the instructor did not speak clearly, ITRS got a mean score of 2.67 with SD of 1.12 interpreted as neutral while NITRS got a mean score of 3.91 with SD of 1.33 interpreted as agree. The instructor did not provide clear explanations, ITRS yielded a mean score of 2.78 with SD of 1.20 interpreted as neutral; on the other hand, NITRS yielded a mean score of 2.40 with SD of 1.35 interpreted as disagree. On the instructor did not consider extended deadlines, ITRS gained a mean score of 2.78 with SD of 1.09 interpreted as neutral while NITRS got a mean score of 2.67 with SD of 1.40 interpreted as neutral.

According to the research, NITRS respondents have difficulty understanding the lesson since their instructor did not talk in a manner that was appropriate for the situation. Respondents even claimed that some professors speak so quickly and utilize unfamiliar terminologies that they are unable to understand what they are trying to communicate. They also have difficulty understanding some computer words because they come from different strands. The ITRS, on the other hand, does not agree or disagree with the provided rationale; similarly, the instructor failed to present a clear discussion with which the NITRS disagrees. Teachers with poor communication skills may cause poor academic performance of students and lead to unstable professional life after school. Good communication minimizes the potential of unkind feeling during the process of teaching and learning. For a teacher, it is very pertinent to have good communication skills to create good classroom environment for effective teacher-student interaction to promote effective learning by students and acquisition of desired professional goals. Good communication is not only needed for effective teaching and learning, but it is also very important in the effectiveness of every human concern in life [22].

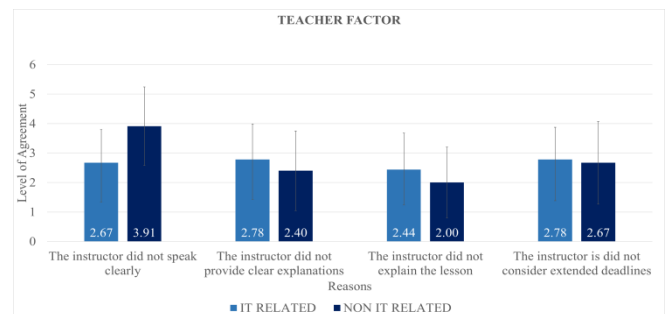


Fig. 4. Reasons for academic deficiencies in terms of teacher factor.

3) *Learning environment*: Fig. 5 shows the reasons for the academic deficiencies in terms of the learning environment. In terms of the internet connection being poor, ITRS obtained a mean score of 3.33 interpreted as neutral on the other hand NITRS received a mean score of 3.67 with SD of 1.54 interpreted as agree. On the learning area was un-comfortable, ITRS got a mean score of 3.33 with SD of 1.22 while NITRS yielded a mean score of 3.40 with SD of 1.24 both interpreted as agree.

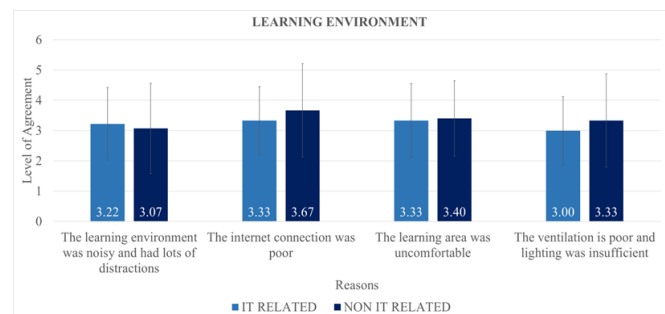


Fig. 5. Reasons for academic deficiencies in terms of learning environment.



Because of the unreliable internet connection, NITRS respondents find it difficult to complete a course satisfactorily, according to the data. Students rely heavily on internet connections in the new normal; from online classroom discussion to assignment submission, a poor internet connection prevents them from joining the class and completing duties. They also underline that being in an uncomfortable learning environment does not help them perform well, and that some people feel uncomfortable when they are being observed doing an activity or participating in a class discussion. Though ITRS does not agree or disagree with either of the above reasons, it is one of their group's highest mean scores, indicating that they are thinking about the reasons for their weaknesses. According to Torres-Diaz, students who do not have access to the internet are more likely to perform poorly on their examinations in comparison to students who frequently utilize the online educational resources. People who participate in interactive activities with their classmates and professors, as well as those who use online resources appropriately for their schoolwork, have a tendency to have better academic achievement overall [23].

4) *Motivation and attitude*: Fig. 6 shows the reasons for the academic deficiencies in terms of the learning environment. On lack of motivation, ITRS gained a mean score of 3.11 with SD of 1.17 while NITRS received a mean score of 3.13 with SD of 1.13 both interpreted as neutral. In terms of failure to do tasks and homework, ITRS obtained a mean score of 2.78 with SD of 1.09 interpreted as neutral on the other hand NITRS yielded a mean score of 3.87 with SD of 1.06 interpreted as agree.

The data shows that NITRS respondents, despite the fact that they did not dislike the courses, accept the fact that they failed to complete the academic task as the reason for their negative academic performance. However, ITRS respondents did not agree or disagree. Furthermore, although being neutral, the highest mean of the latter group lacks enthusiasm to perform as expected and continue with their academic responsibilities.

They lose interest in what they are doing because they see no reason to finish what they have started. Sometimes students lack motivation, so they become apathetic [24]. When students are unmotivated, they might feel that academic success doesn't matter or that they will never achieve it. Students who lack motivation might have experienced a good deal of failure early on in their education and feel there is no point in trying any longer. These students need to find a "why" when it comes to academic success, a reason that will motivate them to achieve their goals. A motivation can come from a career goal, a desire for a future accomplishment, the hope to be financially stable, or even the desire to give back to the community or family members.

5) *Personal issues*: Fig. 7 presents the reasons for academic deficiencies in terms of personal issues. On

procrastinating or delayed doing tasks, ITRS got a mean score of 3.11 with SD of 0.93 interpreted as neutral on the other hand NITRS received a mean score of 3.93 with SD of 1.10 interpreted as agree. Lacking confidence in participating in class, ITRS got a mean score of 3.11 with SD of 1.45 interpreted as neutral while NITRS gained a mean score of 2.80 with SD of 1.37 interpreted as neutral.

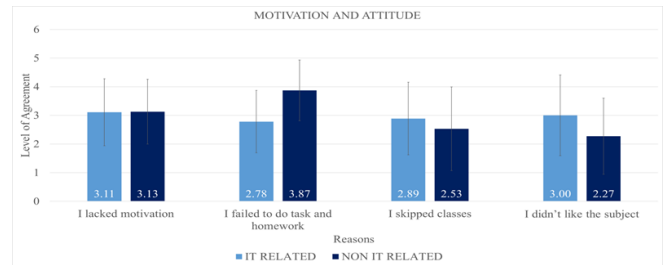


Fig. 6. Reasons for academic deficiencies in terms of motivation and attitude.

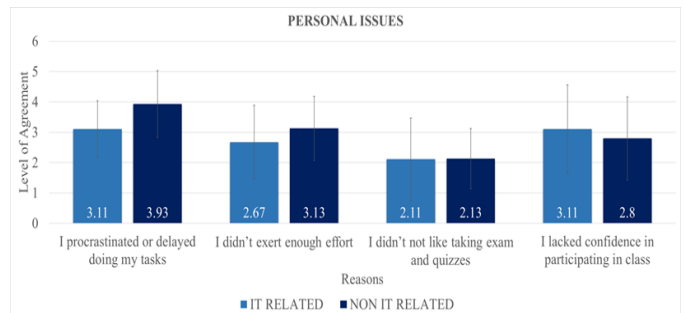


Fig. 7. Reasons for academic deficiencies in terms of personal issues.

According to the research, NITRS respondents have a habit of not completing projects on time. Respondents also admitted to making excuses for not doing their work and deferring chores like schoolwork or preparation for a test by doing anything else instead. ITRS responses do not agree or disagree with the stated personal concerns, but the highest mean in their category agrees with NITRS, and their lack of confidence in engaging in class may be a cause in their academic deficits. Late assignment submission indicates poor quality pupils in terms of performance. Their drop-out percentage (33.33%) is greater than the rate of on-time submissions (5.23%). Their request for an extension on the due date of their tasks is most likely due to poor time management rather than a desire to enhance performance. Students who submit their assignments on time perform better than those who submit their assignments late [25].

### B. Agreement on the Matching of Strand to College Program of Choice as Perceived by the Students

With the above-mentioned reasons, the researcher further investigates if their SHS strand has equipped them with information and skills in preparation for their higher education, and how linked the respondent's strand was to their current program.

TABLE II. THE MATCH OF STRAND TO COLLEGE PROGRAM OF CHOICE AS PERCEIVED BY THE RESPONDENTS

INDICATOR	IT RELATED			NON IT RELATED		
	Mean	SD	INT	Mean	SD	INT
The SHS strand chosen specialization is aligned to the college program of choice.	4.28	0.92	SA	4.00	1.12	A
The specialization subject taken in SHS were related to the courses taken in college	4.21	1.01	SA	4.09	1.01	A
Some knowledge contents needed in college courses are introduced in SHS subjects	4.13	1.06	A	4.09	0.84	A
Some pertinent skills needed in college are taught and developed in SHS	4.24	0.87	SA	4.09	0.98	A
The SHS curriculum meets the college program requirements	4.10	0.94	A	4.09	0.98	A
General Mean and SD	4.19	0.96	A	4.07	0.99	A

Interpretation Legend: SA-Strongly Agree A-Agree N-Neutral D-Disagree SDA-Strongly Disagree

Table II presents the matching of SHS strands based on the perspective of the IT respondents. The result show that in Table I, in terms of the SHS strand chosen specialization is aligned to the college program of choice, ITRS got a mean score of 4.38 with Standard Deviation (SD) of 0.92 and interpreted as strongly agree while the NITRS got a mean score of 4.00 with SD of 1.12 and interpreted as agree. On specialization subjects taken in SHS were related to the course taken in college ITRS got a mean score of 4.21 with SD of 1.01 while the NITRS only agreed with 4.09 mean score with 1.01 SD. Some knowledge content needed in college is taught and developed in SHS and their curriculum meets the college program requirements, ITRS got a mean score of 4.13 and 4.10 with SD of 1.06 and 0.94 while NITRS got a mean score of 4.09 and 0.84 and 0.98 SD, both interpreted as agree. On some pertinent skills needed in college are taught and developed in SHS, ITRS got a mean score of 4.24 with SD of 0.87 and interpreted strongly agree while the NITRS got a mean score of 4.09 with SD of 0.98 interpreted as agree. ITRS on general mean and SD got a mean score of 4.19 and 0.96 while NITRS got a mean score of 4.07 with SD of 0.99 both interpreted as agree.

The data demonstrates that the SHS strand courses, particularly those related to IT; match the courses offered by the IT program. The IT related strand, in contrast to the non-IT-related strand, strongly suggests it is needed in college courses, and the SHS curriculum fits some but not all of the college program standards. College program curriculum is an important factor in a student’s learning development. The structures, standards, and performance indicators of the secondary and tertiary curriculum influence success in college. The usage of a vertically-aligned curriculum is beneficial to the students, although it needs systematic effort. Specialized pre-college courses gear students with advanced knowledge and make them excellent in their chosen fields [26]. During Senior High School, the students will gain the mandatory skills needed for their future professions that will help them do their work in a better way.

C. Correlation between SHS GWA and College GWA

After matching, to further tests the connection between SHS strand to their chosen college program the researcher correlates the SHS GWA to the College GWA to discover if the SHS Strand influences their college academic achievement.

1) Correlation between IT-Related strand and college GWA: Fig. 8 shows the correlation between the SHS GWA of the IT-related strand and their college academic performance.

The data reveals that although technically a positive correlation with  $r = 0.4115$ , the relationship between variables is weak. The p-value is 0.026567. The result is significant at  $p < 0.05$ . By normal standards, the association between the two variables would be considered statistically significant.

2) Correlation between Non-IT related-strand and college GWA: Fig. 9 shows the correlation between the SHS GWA of the Non-IT-related strand and their college academic performance. The data reveals that although technically a negative correlation, the relationship between variables is only weak with  $r = -0.114$ . The P-Value is 0.527587. The result is not significant at  $p < 0.05$ . The association between the two variables would not be considered statistically significant.

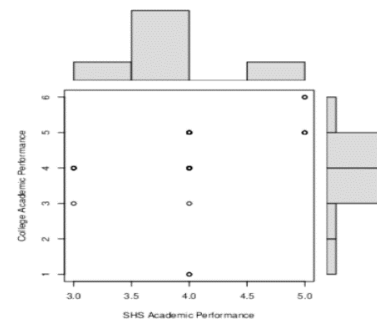


Fig. 8. Correlation between IT-Related strand and college GWA.

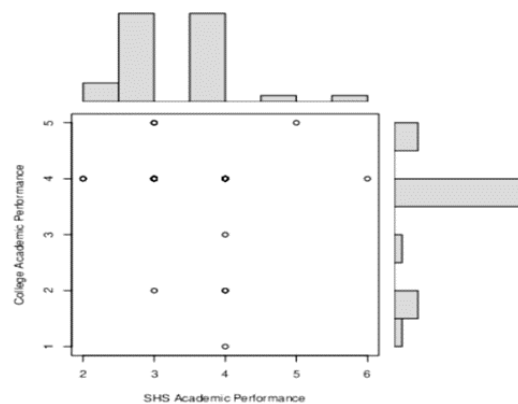


Fig. 9. Correlation between Non-IT related-strand and college GWA.

According to the research, regardless of the strand that IT students choose, it has no effect on their academic performance; nevertheless, enrolling in a related strand has a significant benefit over enrolling in a non-related strand. ITRS respondents also stated that it is simpler for them to grasp the lesson since they have prior knowledge on specific courses. Because of the information they gained in their SHS strand, they are also confident in their ability to contribute and interact in class. In the study of Alipio [27], academic achievement was affected by SHS strand and academic adjustment. Moderation effects between academic adjustment and SHS strand suggest that STEM strand students performed better than those in other SHS strands.

**D. Agreement on How Beneficial the SHS Strand to Academic Success in College**

The study also determines how beneficial the SHS strand is to the academic performance of the respondents' program of choice. This would determine whether or not the learning from their strand was beneficial to their learning acquisition in college. The study also determines how beneficial the SHS strand is to the academic performance of the respondents' program of choice. This would determine whether or not the learning from their strand was beneficial to their learning acquisition in college.

Table III presents how their SHS strand helps their academic performance in college. ITRS perception on SHS strand helps them accomplish their class activities and projects in college obtained a mean score of 4.38 with SD of 0.68 considered as strongly agree while NITRS perception obtained a mean score of 4.15 with SD of 0.87 understood as agree. On SHS strand helps them get high scores in quizzes in college obtained a mean score of 4.07 with SD of 0.70 defined as agree while NITRS obtained a mean score of 4.15 with SD of 0.94 interpreted as agree. ITRS got a general mean of 4.22 and general SD of 0.70 interpreted as strongly agree while NITRS got a general mean of 4.07 with SD of 1.00 and interpreted as agree.

In comparison to the NITRS, the data shows that respondents in the IT-related strand strongly believe that their chosen stand aided their academic achievement in college, particularly in completing class tasks and projects. Some ITRS

respondents stated that some of the college class activities were already presented in their SHS strand, and that the knowledge and abilities required to complete the work had already been gained. Despite the confidence in easy understanding of the course or lesson and task performances, contrary to NITRS experiences due to the mismatch of strand taken in SHS, ITRS cannot confidently believe that their strand was highly effective in getting good grades and test scores because the technique and modality of delivery in higher education are different. However, when compared to non-related strands, aligned strands provide a significant advantage in academic achievement, particularly in the acquisition of knowledge in IT professional courses. According to Lumboy [28] the level of difficulty experienced by the respondents on their college subjects are highly related to the strand they have taken in senior high school. Those who are graduates of STEM excelled over the other as evident in their college academics.

**E. Recommendation to Reduce Academic Deficiencies**

Fig. 10 presents the recommendation to reduce academic deficiencies as perceived by the respondents. The researcher uses thematic analysis to analyze the result. Based on the respondents' recommendation, the analysis produced six (6) themes which included time management, positive attitude and mindset, instructor, requirement completion, study habits, and resumption of classes. Among the themes, requirement completion got the highest percentage of 38%, followed by positive attitude with 25%, study habits and instructor both received 13% while time management is 8% and resumption of classes with the lowest percentage of 4%.

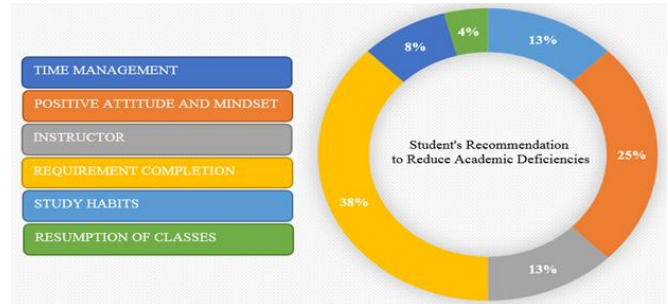


Fig. 10. Recommendation to reduce academic deficiencies.

TABLE III. AGREEMENT ON HOW BENEFICIAL THE SHS STRAND IS TO ACADEMIC SUCCESS IN COLLEGE

INDICATOR	IT RELATED			NON-IT RELATED		
	Mean	SD	INT	Mean	SD	INT
My SHS strand helps me understand the course/lesson in college easily.	4.31	0.66	SA	3.91	1.13	A
My SHS strand helps me accomplish the class activities and projects in college.	4.38	0.68	SA	4.15	0.87	A
My SHS strand gives me confidence in participating in class discussions in college.	4.31	0.66	SA	4.18	0.85	A
My SHS strand helps me perform the class laboratories.	4.20	0.77	SA	3.91	1.2	A
My SHS strand helps me achieve high grades in college	4.03	0.73	A	4.12	1.02	A
My SHS strand help me get high score in quizzes in college	4.07	0.70	A	4.15	0.94	A
General Mean and SD	4.22	0.70	SA	4.07	1.00	A

Interpretation Legend: SA-Strongly Agree A-Agree N-Neutral D-Disagree SDA-Strongly Disagree

1) *Time management*: The necessity of time management in obtaining academic success was underlined by the NITRS respondents in particular. Respondents admitted to not prioritizing their academics as also shown in Fig. 7 and suggested time management as a solution. Knowing what to prioritize, giving time to difficult topic courses, and efficiently organizing their schedule will all help them overcome their academic deficiencies. Some of the responses under this theme include: “In my situation, I had a lot of obligations. Most of the time I can’t manage my time. I need to work to pay my enrollment fee so most of the time I skip classes. My recommendation is time management to avoid failed/INC”, “Give time to the subjects where I am having difficulty”, and “Plan your time well. Do your assignments early enough”.

2) *Positive attitude and mindset*: The majority of NITRS respondents clearly indicate that having a positive attitude and mindset would be extremely beneficial in overcoming academic difficulties. They go on to say that taking a real interest in the course, learning from mistakes, being open to new learning and setting appropriate goals are all things that will help them succeed in school. Negative thinking, according to the respondents, just causes stress and discourages students from pursuing academic success. Some of the responses under this theme include: “Keep a positive attitude. Find genuine interest in the topic. Think about the class and the workload positively”, “Just don’t think of anything that is not related to the very goal of education or the purpose of having a school, especially if those thoughts are just making us stress”, “We must stop pointing and complaining about our conditions as a student. Instead, make solutions to those problems and not excuses, so that we can move on”, and “Learn from mistakes. First of all, you must practice self-criticism and try to analyze the reasons that have caused the failure.”

3) *Instructor*: Instructors play a significant part in every student's academic career, and the instructor's impact can sometimes determine a student's success or failure. Fig. 4 shows that the NITRS respondents have difficulties understanding the lesson because the instructor does not talk properly, and they propose that they convey the discussion in a clear manner. The respondents also recommend that instructors should demonstrate sympathy, be more considerate of students' situations, especially when it comes to submitting activities, and encourage them to pursue their academic goals. In the same way, double-check every submission. Students are motivated to succeed academically when their instructor has a positive view-point and understands their circumstances. Some of the responses under this theme include: “Help the student whenever possible. Ask how you can help. Encourage them and don’t give up on them”, and “I recommend to all teachers to make sure to double check the records of their students.”

4) *Requirement completion*: To avoid academic deficiencies, the majority of respondents recommend meeting all academic prerequisites. Fig. 6 shows NITRS indicating that the reason for their poor academic performance is that they did not complete the task and homework. The respondents then

recommend that students complete all class activities and deliver all required output. The likelihood of passing the course and program increases with a complete submission. Some of the responses under this theme include: “Do all task/project/take quizzes/exam and always participate in online class”, “Try to meet each subjects needed output”, “For me my recommendation to avoid inc is to comply all the requirements, don’t miss out the activity and study more”.

5) *Study habit*: Respondents advise practicing reviewing prior to attending class and taking exams. Additionally, students should concentrate on their studies in order to reduce academic deficiencies. Proper study habits help to improve academic standing, boost confidence in class participation, and achieve high grades on exams. Some of the responses under this theme include: “Focus on the study”, “Review”, and “Don’t be lazy all the time”.

6) *Resumption of classes*: The pandemic has changed the mode of delivery of classes, and some students have been affected and have not fully recovered from the abrupt change in delivery. There were students who excelled in face-to-face delivery but struggled to deliver successfully in online mode, encountering difficulties in meeting the required output and failing to join the class due to a variety of factors. In lieu of the difficulties encountered, a respondent from ITRS has suggested that the face-to-face mode of delivery be resumed to ensure quality learning and that they could truly understand the discussion and avoid academic deficiencies. A response under this theme: “I recommend f2f classes because I think implementing f2f classes will help students to be motivated and has an energy to do things on time and the communication between the teacher and students will really help a lot in the performance of the students.”

#### F. Model for Reducing Academic Deficiencies

Based on the identified causes of academic deficiencies and the respondents' recommendations to avoid them, the researcher developed a model that could potentially reduce academic deficiencies. The model (see Fig. 11) is made up of two (2) components: internal and external, as well as students at the center who will execute and carry out the given components. Internal factors include effective time management, positive attitude and mindset, prompt/punctual requirement completion, and good study habits. Furthermore, in order for the student to remain motivated to achieve academic success, these external factors should be considered; competent and student-friendly instructors, stable, and strong, and accessible internet connections, conducive learning environment, relevant available resources and facilities, adaptation of limited face-to-face or hybrid classes and lastly alignment of SHS strand to college program of choice.

##### 1) Internal factor

a) *Effective time management*: By carefully managing time through meticulous preparation of every second of every day, the student may regulate their schedule so that their time is spent efficiently and they can finally attain their objective. According to Vences Cyril [29], having these skills gives

students the ability to plan ahead and prioritize upcoming assignments and events. This is an important factor in keeping students organized and avoiding procrastination, and ultimately leads to academic success. Being efficient with time may also lead to the discovery of limitless possibilities to focus on oneself and one's mental health, resulting in a right mentality. Ahmad, Batool, & Hussain Ch [30] also mentioned that time management is important in enhancing learners' performance and accomplishments. It is a talent to manage time, and every student must be familiar with and command of this skill in order to get better outcomes. A student can only survive if he or she has good time management skills.

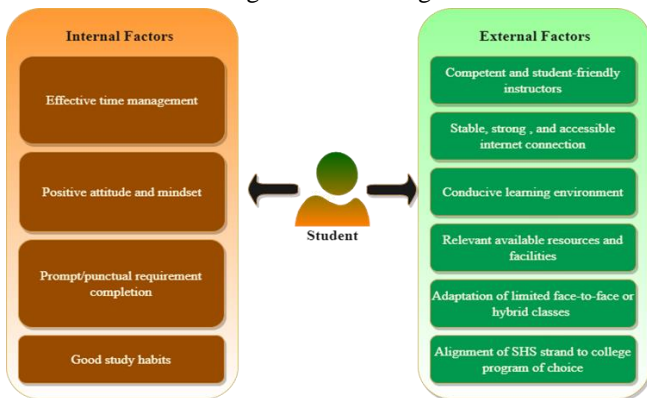


Fig. 11. Model for reducing academic deficiencies.

*b) Positive attitude and mindset:* Having a positive mentality reduces negative thinking, and they will be more focused on studying and executing academic tasks, as well as driven to complete them on time. An article from Overcoming Fear of Academic Failure: Reasons Why Student Fail vs Thrive stated that the ability to learn from failure and keep working toward your goals anyway comes from adopting a resilience mindset. This takes practice, but you can train yourself to understand that failure is part of life and then be willing to keep trying anyway. An important aspect of resilience is the desire to learn from mistakes. Even if you fail, knowing that you can find a solution and work toward your goals in new ways should help you overcome failure [31].

*c) Prompt/punctual requirement completion:* In order to be a successful student, it is essential to complete assignments on time. Meeting time schedule demonstrates that the student can efficiently manage their time and priorities. Numerous activities are evaluated, and timely completion of assignments is essential for receiving excellent scores. By completing assignments on time, students are able to avoid late submission penalties and keep excellent scores in their courses. There are a numerous of benefits associated with doing your assignment in a timely manner. These would include learning how to be more disciplined and independent, improving your ability to manage your time more effectively, which would give you more flexibility to pursue other interests, and increasing the amount of information that you have obtained. By having a great understanding are more likely to achieve a better grade [32].

*d) Good study habits:* There is a substantial association between students' study habits and academic achievement. The study habits of students were used to predict their achievement. The use of study strategies and habits can help students improve their academic performance. Study habits are the reason for the development of academic achievement. [33] [34] [35] Students who have a wide range of self-management skills can deal with long and uncertain deadlines by breaking far-off goals into smaller, closer goals on their own. Students who don't have these skills, on the other hand, would benefit from structures that set sub-goals with deadlines. For example, having students turn in an outline for a paper after the first third of the semester, the first draft after the second third, and the final draft at the end of the semester helps break down a big goal into smaller, more manageable goals. In an ideal world, this scaffolding of self-regulated learning and writing could be used as a model for long-term projects in the future. In general, making goals closer (e.g., with sub-goals) may help the student do better in school and stop putting things off. [36].

## 2) External factor

*a) Competent and student-friendly instructor:* Students' performance is also affected by outside forces that are beyond their control. Because of the diverse worries of the student, having competent and sympathetic instructors will assist the student perform well in class. When the learner thinks that they are being understood, they feel comfortable and motivated to fulfill the work provided. Naz [37] mentioned that for quality education the competent teachers are necessary. Teacher's professional competence includes knowledge and understanding of children and their learning, subject knowledge, curriculum, the education system and the teacher's role. Academic performance has a highly substantial and positively correlated relationship with teachers' levels of expertise, grasp, and clarity on the topics they teach. In addition, aspects of a teacher's demeanor such as engagement, pace, the overall attitude of the instructor, openness, and rapport have a substantial bearing on students' cumulative grade point averages [38] [39].

*b) Stable, strong, and accessible internet connection:* Because of the new normal delivery it is also required to have a solid internet connection in order for the students to join classes, submit requirements and acquire relevant information. Dogniez [40] stated that improved connectivity and the vast learning resources that are available over the Internet can be harnessed to advance access and quality of education. A reliable internet connection offers a lot of convenience and allows us to complete tasks with much less hassle, whether that's downloading a document or making a video call to a work colleague or a loved one. With a good internet connection, we can work, educate ourselves, entertain ourselves and save money.

*c) Conducive learning environment:* A suitable setting is a necessary for learning. Comfortable learning place with minimal destruction helps the student comprehend and grasp the information being taught. In addition, having restricted, substandard and run-down resources and facilities also

hampers student progress, thus it proposes that students should have suitable available learning materials and facilities. This will bring about an enjoyable, exciting and meaningful process, whereby students anticipate attending courses. Students will also feel more at ease learning in a classroom atmosphere and would be able to concentrate more intensively on understanding the subjects at hand. It is supported by Chongui [41] stating conducive environment for learning, such as one with comfortable class-room set-ups, relevant use of teaching materials and interesting classroom activities, will further motivate the students to learn.

*d) Relevant available resources and facilities:* Students' academic success is greatly influenced by the quality of school facilities, and inadequate facilities result in low performance. [42] The availability and proper usage of school physical amenities have a key role in improving students' academic achievement, whereas the lack of such facilities may lead to students' poor academic performance. If physical facilities are accessible and utilized sensibly to fulfill the demands of the students, this would definitely increase students' enthusiasm in studying and lead to superior performance. [43].

*e) Adaptation of limited face-to-face or hybrid classes:* Due to the COVID-19 pandemic, the academic delivery method switched from in-person to online. Some students adapted the new method with ease, but others did not, which affected their academic achievement. There are students that succeed and do better when there is face-to-face interaction, particularly in skills-based courses that need laboratory activities. The researchers chose to incorporate limited face-to-face or hybrid classes in the model in order to meet the demands of students for laboratory activities in light of the ongoing pandemic threat and the contemporary method of instruction.

Adopting hybrid course delivery, which could provide a solution to ensure that Bioscience students receive practical laboratory experience and face-to-face contact to stay motivated and benefit from the on-campus facilities and support, while also giving students some of the flexibility provided by remote study. Student demographics and digital equality must be taken into account in the present competitive higher education market, where student retention is crucial, to ensure that the right strategy is used to serve all students.[44] Kemp [45] also mentioned that with a blended learning strategy, teachers can assign a variety of in-person and online activities to their students. Allowing students to choose how to finish particular task types may be an effective strategy to help more students feel in charge of their own learning and to help them fit learning more readily into their other responsibilities at home and at work.

*f) Alignment of SHS strand to college program of choice:* Students' senior high school track or strand has an impact on their academic success and self-regulated learning. The students' senior high school experience has better prepared them for college, especially in the nursing program. Graduates of senior high school's academic STEM program seem better equipped to enroll in the nursing program, better able to handle the rigors of the program, and more likely to

succeed in nursing school. [46] The model also implies that it would be more advantageous and students would have easier time learning if the strands are connected to the curriculum or perform a bridge program.

Al-Muslimawi and Hamid [47] found external and internal factors that influence students' academic performance in their study. The study found that learning facilities and technology had the greatest impact on students' academic performance, while extracurricular activities can improve academic performance and college life. Competence and aptitude are also excellent indicators of academic success. Family issues might negatively impact students' performance. In addition to socioeconomic position and other factors, work and finances affect students' academic achievement. Social and other issues can slightly affect quality performance, but they are easier to overcome. Moreover, class schedule, size, and atmosphere can negatively impact learning. And, most students' academic performance is affected by text books and test procedures.

In the related studies, significant similarities were discovered, which validates some factors in the researchers' designed model. The researchers, on the other hand, included additional variables that might potentially improve academic performance and fill the gap of the previous studies. Existing studies have identified a number of factors and potential solutions, but the researchers opted to build an approach that concentrated on reducing academic deficiencies. The developed model is a blend of conventional and contemporary solutions that also takes into account the changes in academic mode of delivery caused by the COVID-19 pandemic. And these modifications affected the academic achievement of the students in some way. In addition, because students in this study had an additional year of high school, the researchers investigated whether or not this factored into their academic success in college.

#### IV. CONCLUSION

The researcher discovers that the Non-IT-Related strand, in comparison to the IT-Related strand, has a number of issues in terms of academic performance for a variety of reasons. They emphasize that because of the strand misalignment, they find it difficult to understand the courses in the program chosen. It is also believed that instructors who do not communicate clearly have an impact on students' ability to absorb knowledge, and that students who have an unreliable internet connection have a harder time performing effectively in class. Respondents acknowledged not finishing academic tasks and not submitting them on time, resulting in failed and incomplete remarks. In addition, both the IT-related and non-IT related strands are in agreement that their SHS strands are related to the program that they intend to study in college; however, the IT-Related strand is the one that believes their strand is more beneficial to their academic success in college than the non-IT related strand. The IT-related strand was determined to be statistically significant at a level of  $p < 0.05$ , which indicates that enrolling in the IT-Related strand offers a considerable advantage over the non-IT related strand. As a result, the researchers constructed a model that, with the help of the given reasons and recommendations from the respondents, can potentially eliminate academic inadequacies.

In order to effectively address and give a solution to the problem that has been recognized, the model takes into account both internal and external factors. It is recommended that, with regard to the internal variables, effective time management, a positive attitude and mindset, prompt and punctual completion of requirements, and good study habits be utilized. While for the external factors, the model includes competent and student-friendly instructors, a stable, strong, and accessible internet connection, a conducive learning environment, relevant available resources and facilities, adaptation of limited face-to-face or hybrid classes, and alignment of SHS strand to college program of choice.

The approach that was formulated might be used as a guide by the institution, particularly in the formulation and process of the admission policy to avoid or address misalignment of strand to the college program of choice, evaluation of teaching professionals, and acquisition of education infrastructure. In addition to that, this may also be of use to the Information Technology Department in terms of assessing their curriculum, the approach utilized by teachers, and the monitoring of their students in conjunction with the Guidance Office. Furthermore, this model may be of assistance to the government education sector, particularly CHED and DepEd, in the process of reevaluating the policies and guidelines, particularly as they pertain to the implementation of the K to 12 or Senior High School program and how this program affects academic performance in college.

The investigation was conducted on the IT Department, but the researchers recommend looking into other programs. Additionally, to perform in-depth study on the alignment of the curriculum between senior high schools and higher education institutions, as well as how it contributes to other possible causes of academic inadequacies.

#### ACKNOWLEDGMENT

Above all the authors thank and praise Almighty God for making everything possible. Sincere appreciation to the respondents of the study, to Dr. Thelma Palaoag, Dr. Regina Sarte and Dr. Florivel Villacorta for the humble ideas and support. The panel of experts for the constructive inputs and precious time for the improvement of the study. And lastly, authors' family and colleagues for the motivation and inspiration.

#### REFERENCES

[1] L. R. P. Magdadaro, (2020). Passion-based vs. Practical-based preference of strand in senior high school. *Int J Acad Res Bus Soc Sci*. 2020;10.

[2] S. Masud, S. H. Mufarrih, N. Qaisar , F. Khan, S. Khan and M. N. Khan, "Academic Performance in Adolescent Students: The Role of Parenting Styles and Socio-Demographic Factors – A Cross Sectional Study From Peshawar, Pakistan," 08 November 2019.

[3] R. Rono, 2013. Factors Affecting Pupils' Performance in Public Primary Schools at Kenya Certificate of Primary Education Examination (Kcpe) in Emgwen Division, Nandi District, KENYA. <http://erepository.uonbi.ac.ke/bitstream/handle/11295/52949/ABSTRACT.pdf?sequence=43>. (Accessed 6 January 2018). Date accessed: February 5, 2022.

[4] A. A. Hayat, K. Shateri, M. Amini and N. Shokrpour, "Relationships between academic self-efficacy, learning-related emotions, and

metacognitive learning strategies with academic performance in medical students: a structural equation model," 17 March 2020.

[5] S. Ali, Z. Haider, F. Munir, H. Khan, & A. Ahmed, (2013). Factors Contributing to the Students' Academic Performance. A case study of Islamia University, Sub-Campus. *Science and Education Journal*, Vol. 1 issue8 pp1-10.

[6] O. T. Olufemi, A. A. Adediran and W. Oyediran, "Factors affecting students's academic performance in Colleges of Education in Southwest, Nigeria," *British Journal of Education*, vol. Vol.6, no. No.10, pp. pp.43-56, October 2018.

[7] S. Masud, S. H. Mufarrih, N. Qaisar , F. Khan, S. Khan and M. N. Khan, "Academic Performance in Adolescent Students: The Role of Parenting Styles and Socio-Demographic Factors – A Cross Sectional Study From Peshawar, Pakistan," 08 November 2019.

[8] M. Clercq, B. Galand, S. Dupont, & M. Frenay, (2013). Achievement among first-year university students: an integrated and contextualised approach. *European Journal of Psychology of Education*, 3(28), 641-662.

[9] O. T. Olufemi, A. A. Adediran and W. Oyediran, "Factors affecting students's academic performance in Colleges of Education in Southwest , Nigeria," *British Journal of Education*, vol. Vol.6, no. No.10, pp. pp.43-56., October 2018.

[10] E. A. Brew, B. Nketiah and R. Koranteng, "A Literature Review of Academic Performance, an Insight into Factors and their Influences on Academic Outcomes of Students at Senior High Schools," vol. Vol.8, no. No.6, June 2021.

[11] C. A. Quintos, D. G. Caballes, E. M. Gapad, & M. R. Valdez, (2020). Exploring Between SHS Strand and College Course Mismatch: Bridging the Gap Through School Policy on Intensified Career Guidance Program. *Data Mining and Knowledge Engineering*, Vol 12, No 10-12 (2020), 156-161.

[12] N. Maloshonok, & E. Terentev, (2017). The mismatch between student educational expectations and realities: prevalence, causes, and consequences. *European Journal of Higher Education*, 7(4), 356-372. Retrieved from <https://doi.org/10.1080/21568235.2017.1348238>.

[13] J. M. Catedrilla, J. M. De La Cuesta, & M. R. Caguia, (2019). Impact of the STEM Program on Information Technology College Students' Goals: Perspectives from the Philippines. *Proceedings of the 27th International Conference on Computers in Education*. Taiwan: Asia-Pacific Society for Computers in Education.

[14] M. M. Alipio, (2020, March). Academic Adjustment and Performance among Filipino Freshmen College Students in the Health Sciences: Does Senior High School Strand Matter? Davao City, P hilippines.

[15] G. Chirtes, (2010). A case study into the causes of school dropout. *Acta Didactica Napocensia*. 3(4), 2065-1430.

[16] Ş. Şahin, Z. Arseven, & A. Kılıç, (2016, January). Causes of Student Absenteeism and School Dropouts. *International Journal of Instruction*, Vol.9 (No.1).

[17] W. Fan., & C. A. Wolters, (2014). School motivation and high school dropout: The mediating role of educational expectation. *British Journal of Educational Psychology*, 84, 22–39.

[18] S. M. Al-Zoubi and M. A. Bani Younes, "Low Academic Achievement: Causes and Results," *Theory and Practice in Language Studies*, vol. Vol. 5, no. No. 11, pp. pp. 2262-2268, November 2015.

[19] V. Clarke and V. Braun, (2013) Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The Psychologist*, 26(2), 120-123.

[20] "Overcoming Fear of Academic Failure: Reasons Why Students Fail vs Thrive," 20 May 2020. [Online]. Available: <https://www.nshss.org/blog/overcoming-fear-of-academic-failure-reasons-why-students-fail-vs-thrive/>.

[21] R. J. Collie, A. J. Martin, B. Papworth, & P. Ginns, (2016). Students' Interpersonal Relationships, Personal Best (P.B.), Goals, and Engagement. *Learning and Individual Differences*, 45, 65-76. doi:<https://doi.org/10.1016/j.lindif.2015.12.002>.

[22] E. Obilor, (2020). Teacher Factors Influencing Students' Academic Performance in Public Secondary School in Rivers State. *International Academy Journal of Educational Technology and Research*, 28-41.

- [23] Torres-Diaz. (2016). Internet Use and Academic Success in University Students. *Media Education Research Journal*, 61-70.
- [24] "Overcoming Fear of Academic Failure: Reasons Why Students Fail vs Thrive," 20 May 2020. [Online]. Available: <https://www.nshss.org/blog/overcoming-fear-of-academic-failure-reasons-why-students-fail-vs-thrive/>.
- [25] T. Maraseni & G. Cockfield, (2006). Analysis of 'On-time' and 'Late' Assignment. *International Journal of Business & Management Education*, 14(2), 23.
- [26] J. C. B. Jacolbia, S. N. M. Balitaan, J. D. Deloria, M. A. C. Perey, H. M. S. Publico, P. M. R. Reyes and D. L. G. Vicencio, "The Comparison of Senior High School Track and College program references and the Factors Affecting college course decision of Selected ABM students of the Polytechnic University of the Philippines," March 2018.
- [27] M. M. Alipio, (2020, March). Academic Adjustment and Performance among Filipino Freshmen College Students in the Health Sciences: Does Senior High School Strand Matter? *Education and Management*.
- [28] M. Lumboy, "Senior High School Strand Choice: Its Implication to College Academic Performance," *Ascendens Asia Journal of Multidisciplinary Research*, vol. Vol. 3, no. No. 7, 18 December 2019.
- [29] A. Vences Cyril, "Time Management and Academic Achievement of Higher Secondary Students," *Journal on School Educational Technology*, vol. Vol. 10, no. No. 3, 2015.
- [30] S. Ahmad, A. Batool, & A. Hussain Ch, (2019). Path Relationship of Time Management and Academic Achievement of Students in Distance Learning Institutions. *Pakistan Journal of Distance & Online Learning*, V(II), 191-208.
- [31] "Overcoming Fear of Academic Failure: Reasons Why Students Fail vs Thrive," 20 May 2020. [Online]. Available: <https://www.nshss.org/blog/overcoming-fear-of-academic-failure-reasons-why-students-fail-vs-thrive/>.
- [32] Importance of Turning Work On Time and Being On Time. (2021, July 28). Retrieved from WritingBros: <https://writingbros.com/essay-examples/importance-of-turning-work-on-time-and-being-on-time/>.
- [33] M. Rabia, N. Mubarak, H. Tallat & W. Nasir, (2017). A Study on Study Habits and Academic Performance of Students. *International Journal of Asian Social Science*, 7(10), 891-897. doi:10.18488/journal.1.2017.710.891.897.
- [34] N. Alimohamadi, M. Dehghani, S. Saeide & E. Ashtarani, (2018). Relation Study between Study Habit and Academic Performance of Nursing Students in Hamadan. *Pajouhan Scientific Journal*, 16(3), 29-38. doi:10.21859/psj.16.3.29.
- [35] K. Okado, N. Kida, & T. Sakai, (2018). Changes in Study Habits of Chinese Adolescents and Factors Supporting These Habits-Focusing on the Transition Period from Elementary School to Junior High School. *Psychology*, 9(5), 1081-1094. doi:10.4236/psych.2018.95068.
- [36] P. Steel, F. Svartdal, T. Thundiyil, & T. Brothen, (2008). Examining Procrastination Across Multiple Goal Stages: A Longitudinal Study of Temporal Motivation Theory. *Frontiers in Psychology*. doi:10.3389/fpsyg.2018.00327.
- [37] K. Naz, "Effects of teachers' professional competence on students' academic achievements at secondary school level in Muzaffarabad District," 2016.
- [38] M. Rashid, & S. Zanab, (2018). Effects of Teacher's Behavior on Academic Performance of Students.
- [39] P. Riahipour, & A. Dabbaghi, (2014). Iranian EFL Teachers' Perceptions of Traditional, Innovative and. *The Iranian EFL Journal*, 18(2), 268.
- [40] J. Dogniez, "Internet for Education: Key considerations for Advancing Sustainable Development," January 2019. [Online]. Available: <https://www.dotmagazine.online/issues/socially-responsible-digitalization/doteditorial-ethical-standards-for-digital/internet-for-education>.
- [41] L. Chonghui, "Conducive environment a must for Learning," 26 July 2020. [Online]. Available: <https://www.thestar.com.my/news/education/2020/07/26/conducive-environment-a-must-for-learning>.
- [42] F. D Oluremi, & O. O. Olubukola, (2013). Impact of facilities on academic performance of students with special needs in mainstreamed public schools in Southwestern Nigeria. *Journal of Research in Special Educational Needs*. *Journal of Research in Special Educational Needs*, 13(2), 159-167. Retrieved from <https://doi.org/10.1111/j.1471-3802.2011.01228.x>.
- [43] C. O. Akomolafe, & V. O. Adesua, (2016). The Impact of Physical Facilities on Students Level of Motivation and Academic Performance in Senior Secondary Schools in South West Nigeria. *Journal of Education and Practice*, 7(4).
- [44] A. Bashir, S. Bashir, K. Rana, P. Lambert, & A. Vernallis, (2021). Post-COVID-19 Adaptations; the Shifts Towards Online Learning, Hybrid Course Delivery and the Implications for Biosciences Courses in the Higher Education Setting. *Frontiers in Education*. doi:10.3389/educ.2021.711619.
- [45] N. Kemp, (2020). University students' perceived effort and learning in face-to-face and online classes. *Journal of Applied Learning and Teaching*, 3(1). doi:10.37074/jalt.2020.3.s1.14.
- [46] X. G. Malaga, & R. F. Oducao, (2021). Does Senior High School Strand Matter in Nursing Students' Academic Self-Regulated Learning and Academic Performance? *South East Asia Nursing Research*, 3(1). doi:10.26714/seanr.3.1.2.21.1-7.
- [47] I. Al-Muslimmawi, & A. Hamid, (2019). External and Internal Factors Affecting Students' Academic Performance. *The Social Sciences*, 14(4), 155-168.



# AI based Dynamic Prediction Model for Mobile Health Application System

Adari Ramesh<sup>1</sup>, Dr. C K Subbaraya<sup>2</sup>, Dr. G K Ravi Kumar<sup>3</sup>

Research Schaler-Department of Computer Science and Engineering, College of BGS Institute of Technology (BGSIT)

Adichunchanagiri University (ACU), B.G. Nagara, Nagamangala, Karnataka<sup>1</sup>

Registrar, Adichunchanagiri University (ACU), B.G. Nagara, Nagamangala, Karnataka<sup>2</sup>

Department of R&D (CScience&E)-IT Head, Adichunchanagiri University (ACU), B.G. Nagara, Nagamangala, Karnataka<sup>3</sup>

**Abstract**—In recent decades, mobile health (m-health) applications have gained significant attention in the healthcare sector due to their increased support during critical cases like cardiac disease, spinal cord problems, and brain injuries. Also, m-health services are considered more valuable, mainly where facilities are deficient. In addition, it supports wired and advanced wireless technologies for data transmission and communication. In this work, an Artificial Intelligence (AI)-based deep learning model is implemented to predict healthcare data, where the data handling is performed to improve the dynamic prediction performance. It includes the working modules of data collection, normalization, AI-based classification, and decision-making. Here, the m-health data are obtained from the smart devices through the service providers, which comprises the health information related to blood pressure, heart rate, glucose level, etc. The main contribution of this paper is to accurately predict Cardio Vascular Disease (CVD) from the patient dataset stored in cloud using the AI-based m-health system. After obtaining the data, preprocessing can be performed for noise reduction and normalization because prediction performance highly depends on data quality. Consequently, we use the Gorilla Troop Optimization Algorithm (GTOA) to select the most relevant functions for classifier training and testing. Classify his CVD type according to a selected set of features using bidirectional long-term memory (Bi-LSTM). Moreover, the proposed AI-based prediction model's performance is validated and compared using different measures.

**Keywords**—Artificial Intelligence (AI); M-Health System; Data Collection; Cloud Storage; Gorilla Troop Optimization (GTO); Bi-directional Long Short-Term Memory (Bi-LSTM); dynamic prediction

## I. INTRODUCTION

The healthcare industry faces several challenges in diagnosing disease and providing affordable services [1, 2]. Providing patients with the best possible care based on a review of their medical histories, medical decisions, and the variability of their molecular properties is one of the fundamental requirements of any healthcare system. Due to the rapid development of new technologies, these systems have been dealing with several problems with data gathering, information association, data retrieval, and decision-making [3]. Due to logistical constraints caused by mobile phone use in developing nations, many healthcare sectors have significantly transformed. Mobile technology [4] has recently been crucial in several technological domains among

subscribers in practically all countries. The advancement of new technologies and their use in the healthcare industry, known as "m-health" [5, 6] are aided by mobile devices and communications. The general framework of m-health system is shown in Fig. 1. Still, the m-health framework [7, 8] has many challenges and issues, which include the followings:

- Better interpret the data from multiple sources produced by numerous mobile and information sources after 2021.
- Enabling smarter, more personalized behavior change and engaging tools to inspire more meaningful users and patients to improve their health and well-being. The vast amount of health data from 5G mobile health users also needs to be intelligently adapted and transformed. It is essential to accurately estimate the relationship between the genomic data sequence and other medically relevant data.

The interpretation of large medical imaging datasets and other relevant diagnostic/imaging data generated by the latest generation of mobile imaging devices requires reliable, accurate and secure data analysis methods.

For instance, the cardiac arrest is now more likely in the situation where life is currently halted. Patients' health conditions deteriorate when they put off getting medical care because they are worried about contracting a communicable illness. For diagnosis and therapy, accurate predictions are essential. Researchers are constantly creating useful decision assistance systems, but CVD diagnosis is still difficult. Several intelligent tools have been developed based on machine learning and data-driven methodologies to address these issues. These methods have all included connecting many data sources to create a collective understanding for future research and predictive analysis [9, 10]. Several research have demonstrated that the severity of heart diseases may be automatically diagnosed using various machine learning approaches, such as combining numerous classification algorithms and augmentation algorithms to create reliable automated prediction systems. The CVD dataset was utilized in the research to test seven models, including SVM, KNN, LR, DT and various ensemble methods. Moreover, the AI-assisted diagnosis can help doctors make diagnoses more quickly and cheaply, which can help save patients' lives earlier in the process when there is a shortage of medical professionals.

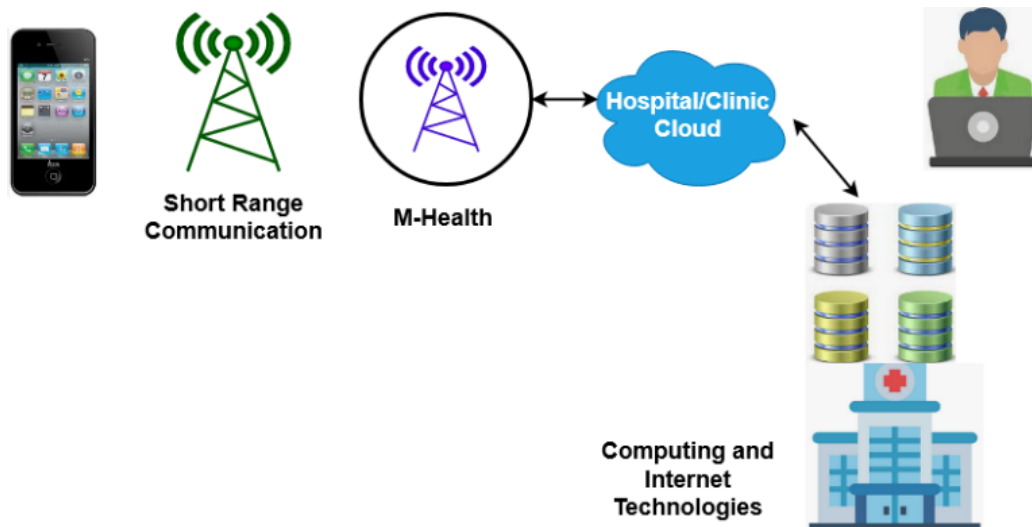


Fig. 1. General framework of M-Health systems.

Furthermore, many prediction-based techniques have been developed for various conditions, including neurological and cardiovascular heart disorders. In m-health systems, patient data is gathered to help with disease tracking, diagnosis, and prompt services, particularly in places where immediate care is complicated. Another online platform with a substantial library of medical content built on multimedia is called E-Medicine. Despite the availability of these new communication channels [11, 12], the technological support for the healthcare industry often needs to meet expectations due to scalability and other technical issues. To make the service more accessible to providers, users, and medical professionals, a more suitable m-health architecture needs to be designed. In addition, all data transmissions must be quick and secure to ensure timely delivery [13]. Security is one of the most crucial elements, particularly in relation to wearable sensors and handheld devices. The major research objectives of this paper are as follows:

- To develop a computationally efficient M-Health framework by using an advanced Artificial Intelligence (AI) mechanism.
- To accurately predict the cardiovascular disease from the obtained patient health information with reduced time and computational complexity.
- To improve the prediction rate of the M-Health systems, a Gorilla Troop Optimization Algorithm (GTOA) mechanism is deployed.
- A bidirectional long-short-term memory (Bi-LSTM) model has been implemented to ensure efficient and accurate disease diagnosis.

A comprehensive analysis was performed using various performance measures to validate and test the results of the proposed GTOA using the Bi-LSTM model.

The following sections make up the remaining parts of this article: The traditional approaches and methods for dynamic disease prediction in m-health systems are reviewed in Section II. In accordance with its primary disease prediction

outcomes, it also analyses the benefits and drawbacks of each technique. The entire description of the proposed approach, including its workflow and extensive illustrations, is provided in Section III. Using a variety of parameters, Section IV compares and validates the results of the proposed and existing disease prediction models. In Section V, the study concluded with its findings, implications, and future scope.

## II. RELATED WORKS

This section gives an exhaustive assessment of the literature on the current AI development approaches. Additionally, it looks at each work's benefits and drawbacks in relation to its performance, core concepts, and processing method.

The most popular machine learning approaches are classification algorithms. Based on their preferred method of learning, machine learning models can be broadly categorized into three categories: reinforcement learning, unsupervised learning, and supervised learning. When a machine learns how to assign labels to each class of data that process is known as classification. Moreover, the ANN, NB, DT, and LR are the most popular machine learning models specifically used in the healthcare applications for disease prediction. To identify patterns and address AI issues, the ANN is constructed using a network of neurons and weighted connections. During training, the random forest generates a number of decision trees and produces the results that receive the most votes. This method is applied to tasks involving classification and regression. NB is a Bayes' rule-based classification system that works under the presumption that all of the features that forecast the ideal value are unbiased. Based on the likelihood that the data belong to a particular class, the Nave Bayes model can determine the class of a given set of data. A straightforward method known as "nearest neighbors" can be used to save all existing cases and categorize new cases based on how similar they are as determined by distance functions.

Istepanian and Anzi [14] utilized a new m-health framework for an intelligent healthcare delivery systems. The purpose of this work was to investigate the relevant big data

issues, and technological innovations for developing an effective m-Health framework. Moreover, we validated different types of data analysis approaches for the M Health system, including descriptive models, diagnostic models, predictive models, and prescriptive models. Based on this research, machine learning and deep learning tools are analyzed to be of great importance and play an important role in the m-Health framework. Mishra et al. [15] pursued a new approach to improve the effectiveness of m-health systems using big data and IoT technologies. This paper targets to construct a new IoT based m-Health framework for disease monitoring and patient empowerment. Also, it intends to minimize the cost and enable an effective utilization of the medical components for proper m-Health management and control. However, the system model of the suggested framework could be difficult to understand, and it failed to demonstrate the efficiency and reliability of this systems. Alotaibi, et al. [16] analyzed the major applications of using AI and big data analytical models for developing a highly proficient m-Health systems. Here, a systematic review is conducted for investigating the different types of AI mechanisms to improve the performance of m-health systems. Moreover, this work used various performance measures to assess the quality of m-health systems that includes interactivity, veracity, usefulness, effectiveness, user satisfaction, and customization. The major drawbacks of this framework are inaccurate system response, various privacy and security problems.

Al-Marridi, et al. [17] employed an AI technique for optimizing the energy efficiency of m-health systems. The purpose of this work is to effectively minimize the latency and jitter by properly providing the resources and services from the cloud data centers. Moreover, it mainly concentrated on the optimal allocation of resources in the m-health systems. For this purpose, the deep learning based neural network algorithm was deployed, which supports to reduce the distortion and maximize the compression ratio of the healthcare framework. The primary advantages of this framework are reduced time consumption, optimized energy consumption, and delay. Elhishi, et al. [18] deployed a mobile health application system for identifying the leukemia cancer, where a straightforward medication management scheme was also used to remind the patients about their schedules. Specifically, it intends to develop a construct a new prototype for identifying the suitable solution with reduced time and cost. In addition, the Convolutional Neural Network (CNN) based deep learning classifier has been utilized to detect the leukemia cancer using the blood film input. Lano, et al. [19] utilized an AI mechanism to assist the m-health application systems. Here, the machine learning model is utilized to analyze the behavior elements of the medical systems. Mendo, et al. [20] provided a thorough analysis of the many machine learning models employed to raise the capability of mobile health systems. We also explored the rise of mobile healthcare devices and applications for reviewing patient health records. Pankaj et al. [21] used a machine learning approach to diagnose diabetes using the M-Health application. Abed et al. [22] introduced an integrated electronic medical system that enables efficient medical monitoring. Here, the machine learning based traffic flow classification model was utilized

for reducing the time delay and traffic during data transmission/communication.

Alhussein, et al. [23] developed a voice pathology detection system with the help of CNN model for mobile health applications. The purpose of this work was to design a smart m-health framework by using the transfer learning mechanism. In addition, various types of features such as mel-frequency cepstral coefficients (MFCC), pitch frequencies, and linear prediction cepstral coefficients (LPCC) are also used. Shaban-Nejad, et al. [24] presented a detailed analysis demonstrating the importance of using AI in personalized healthcare systems. The scope of this paper was to improve the speed, accuracy, and reduce the time in the health systems. Shatte, et al. [25] presented a comprehensive survey for examining various ML methodologies used for developing a computationally proficient healthcare monitoring framework. Moreover, it discussed about the importance of using ML techniques in the field of medical treatment and diagnosis. It includes the mechanisms of active learning, Bayesian network, ensemble learning, regression, KNN, multivariate classification, random forest, linear discriminant analysis, and discriminative dictionary learning. Dargan, et al. [26] presented a comprehensive survey for examining the different types of deep learning mechanisms used in a health application systems. Key factors for using deep learning methodologies are layered or multi-level nonlinear processing and whether learning is guided or unguided. Garcia et al. [27] developed a framework for mental health monitoring systems (MHMS) based on machine learning models. The main consideration of this paper is to ensure the properties of increased privacy, high storage capacity, reliable data transmission, reduced energy consumption, and flexible data labeling. Tian, et al. [28] intended to design a smart healthcare framework with the specific parties of efficiency, convenience, and personalization. This framework includes the major participants of doctors, patients, hospitals and research institutions.

Based on the literature review, the problems associated to the conventional M-Health frameworks are as follows:

- Complex system design.
- Increased error prediction rate.
- Computational burden.
- High time consumption.
- More resource utilization.

Therefore, the proposed work motivates to develop a new M-Health framework for dynamic disease prediction.

### III. PROPOSED METHODOLOGY

This section provides the complete explanation for the proposed m-health systems with the overall workflow and illustrations. The original contribution of this work is to predict the cardiovascular disease according to the patient information collected from the cloud storage by using an advanced AI mechanism. The architectural model of the proposed M-Health framework is shown in Fig. 2. The development of mobile systems, especially for emergency

applications, is usually a product of information and communication technology. Moreover, the emergency applications are highly crucial for the patients who suffered with heart disorders, brain injuries, spinal injuries, and etc. The M-Health services are beneficial, particularly in areas without many medical facilities, where hospitals are dispersed among the population, or where the cost of medical care is high. Also, it is built on new wireless and wired technologies, including cloud computing, Global System for Mobile Communications (GSM), and 4<sup>th</sup> or 5<sup>th</sup> generation technologies. Smart sensors are a feature of M-Health systems, which also have 5G connectivity capabilities linked with Web 2.0, online communities, and cloud - based services. Furthermore, the majority of M-Health-capable sensors and gadgets use low-power Bluetooth and ZigBee technologies to transmit data to other endpoints. The information is then shared to the distant systems or cloud data storage over communication networks for additional processing and decision-making. The modules involved in the proposed M-Health framework are as follows:

- Data collection using smart sensors.
- Data transmission to cloud.
- Deep learning based cardiovascular disease prediction.

Initially, the data collection is performed through smart sensors that are planted inside the human body, which are further connected with the mobile devices or gateways. Then, the smart devices like mobile can be used to enable communication between the patients and healthcare professionals at anywhere. After that the obtained health information is transmitted to the cloud systems for further operations. Furthermore, the deep learning based AI mechanism is applied to accurately predict the cardiovascular disease from the obtained patient health information. As shown in Fig. 3, the proposed AI based disease prediction system comprises the following operations:

- Secure data collection from cloud or data centers.
- Data preprocessing.

- Gorilla Troop Optimization Algorithm (GTOA) for feature selection.
- Bi-directional LSTM based classification.

After obtaining the health data from cloud systems, the data preprocessing is applied to generate the balanced dataset by performing the operations of cleaning, normalization, noise removal and balancing. Since, the raw dataset comprises some un-related attributes or information that may disrupt the accuracy of disease detection. Hence, the data preprocessing is more essential for an accurate disease prediction and diagnosis. After that, an intelligent and computationally efficient GTOA is applied for reducing the dimensionality of the dataset by choosing the most relevant features. It is also helps to improve the disease detection rate and reduce the overall time consumption of the processing system. Finally, the deep learning based AI model, named as, Bi-LSTM has been applied to accurately predict the type of cardiovascular disease according to the set of extracted feature set. The performance of the proposed GTOA using the Bi-LSTM model is validated and tested with various parameters. The primary advantages of the proposed model are as follows: easy to understand, reduced complexity in processing, less time consumption, accurate disease prediction, and computationally proficient.

#### A. Cloud Model

In this framework, the cloud system comprises two layers such as data annotation & analysis, and storage & access layers, in which the storage layer stores the M-Health information obtained from the mobile devices through network service providers. The data is related to the patient's medical information like previous history, glucose level, BMI, heart rate, etc. Typically, the cloud is one of the most suitable and convenient platform for the users to store and retrieve the information. Specifically, the cloud has the major benefits of reduced cost, stability, and efficiency for the M-Health applications. Here, isolation strategies are used to enable secure data storage operations that ensure privacy and confidentiality of patient medical information. The data storage model of the cloud system for the proposed M-Health application system is shown in Fig. 4.

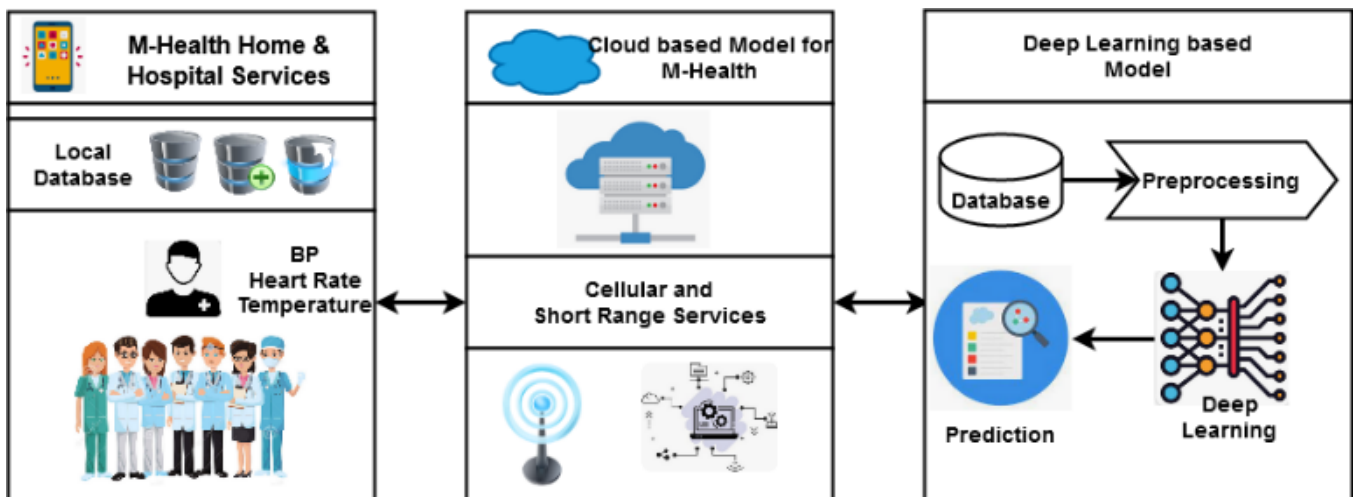


Fig. 2. M-Health framework of the proposed system.

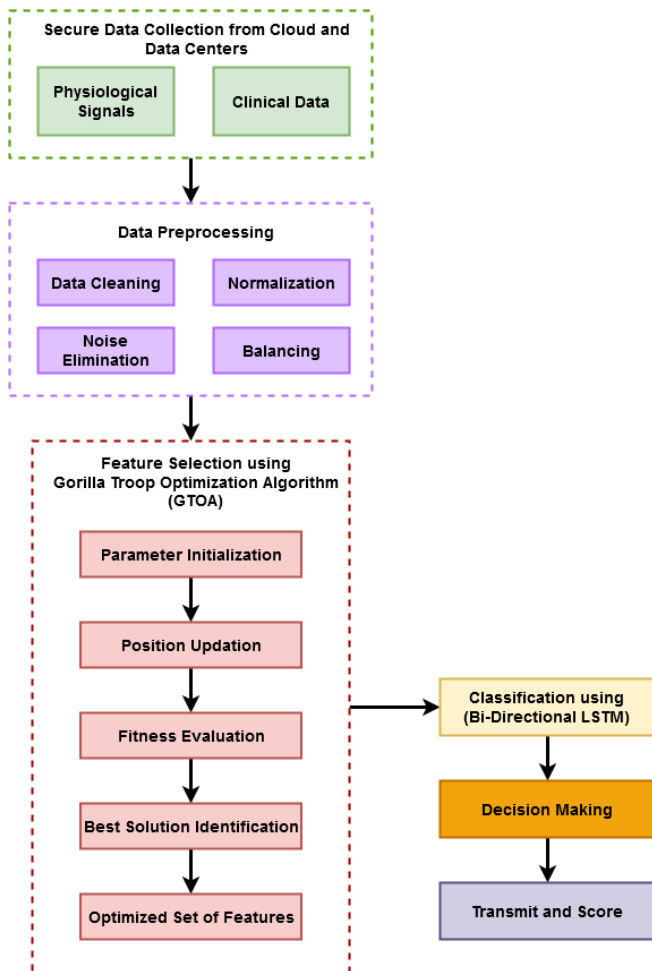


Fig. 3. Flow of the proposed dynamic prediction system.

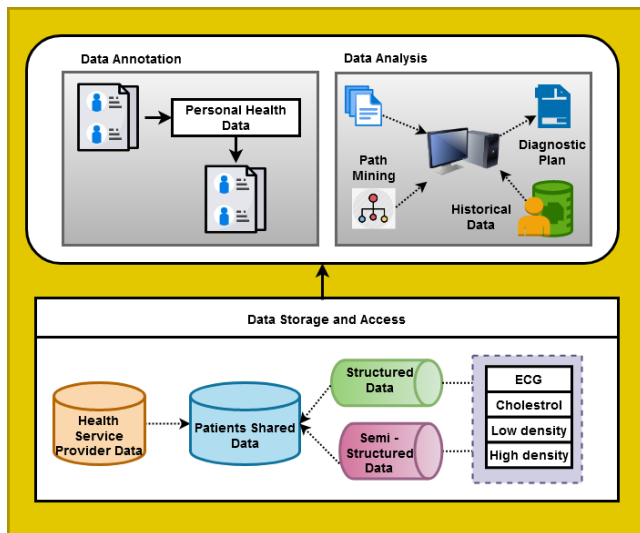


Fig. 4. Cloud data storage model.

### B. Secure Data Collection

The health or medical information about the patients like clinical records, and physiological signals are gathered from the smart sensor devices. Here, the wearable smart devices monitor the patients' data, and transmits the obtained

information to the mobile device. Since, the existing data are gathered from the cloud systems, which comprise the information of patient history, age, BMI, gender, blood pressure, etc. This information is updated by the medical centers, and it can be used to predict that whether the patient is affected by the CVD or not.

### C. Data Preprocessing

After collecting the information, we apply data preprocessing to create a noise-free and balanced dataset for accurate disease prediction and classification. The main goal of this work is to accurately detect CVD disease in order to provide appropriate treatment and recommendations to patients. Data normalization and noise reduction techniques are used to make the data noise-free at this stage. Noise and other environmental influences are present in the raw data that comes from the sensor nodes. Moreover, it includes the preprocessing operations of cleaning, normalization, noise elimination, and data balancing. This technique is used to tune the noisy data into a clean dataset. In other words, anytime data is acquired from various sources, it is done in a primary manner that makes analysis impossible. So that the deep learning model being used can produce better outcomes, data must be presented in the correct format. Pre-processing describes the changes made to our data before the algorithm receives it. The removal of mistakes, inconsistencies, and missing values from the dataset is the focus of the data cleansing/cleansing stage. Various approaches or techniques have been developed to address this issue. The data value missing from the cell in the corresponding column is known as a missing value. In healthcare, missing values may occur due to human error, non-applicability, sensor omission, patient absence from the ventilator due to medical decisions, or patient state unrelated to a particular variable. Moreover, the missing values can be handled by either rejecting them or imputing the missing information. In addition, the data scaling is performed by using the min-max algorithm, which scales the features in the range of [0, 1], and [-1, 1]. It is estimated as follows:

$$D = \frac{d - d_{mn}}{d_{mx} - d_{mn}} \quad (1)$$

Where,  $d_{mn}$  and  $d_{mx}$  are the minimum and maximum values of the dataset. Based on this operation, the balanced and quality improved data is produced as the output, which can be used for an accurate CVD prediction.

### D. Gorilla Troop Optimization Algorithm (GTOA)

After preprocessing, the GTOA is applied to optimally select the features from the balanced dataset for accurately predicting the disease. This technique provides the best solution for setting features to train the classifier. Finding the best viable or desirable solutions to a specific problem that is frequently faced in a variety of fields is referred to as optimization. Multiple factors have contributed to the popularity of meta-heuristics in engineering applications. For instance, they are simple to execute, contain reasonably simple concepts, perform better than local search algorithms, have many different applications, and knowledge of the derivative function is not required. In the existing disease prediction frameworks, various meta-heuristics optimization techniques

are deployed for feature optimization and dataset dimensionality reduction. However, it faces some complications associated with slow convergence, a high number of iterations, complexity in computation, and more time consumption. Hence, the GTOA is utilized in this work, which supports reducing the dimensionality of the preprocessed dataset by selecting the most relevant features used for CVD prediction.

In this technique, the five distinct operators are used to simulate exploration and exploitation optimization procedures based on the behaviors of gorilla, and its phases are shown in Fig. 5. During exploration, three distinct operators have been used such as:

- Migration to unknown place for increasing the capability of exploration.
- Migration to other gorillas for balancing both exploration and exploitation.
- Move in the direction of an identifiable place to increase the capability of GTO.

These operations are mathematically performed as shown in below:

$$AG(h+1) = \begin{cases} (Up_b - Lo_b) \times \alpha + Lo_b & r < d \\ (\beta - K) \times G_r(h) + H \times B & r \geq 0.5 \\ G(i) - K \times (K \times (G(h) - AG_r(h)) + \gamma \times (G(h) - AG_r(h))) & r < 0.5 \end{cases} \quad (2)$$

Where,  $AG(h+1)$  indicates the candidate position vector of gorilla at iteration  $h$ ,  $G(h)$  denotes the current vector of gorilla,  $r, \alpha, \beta, \gamma$  are the random numbers from [0 to 1] that can be updated at each iteration,  $d$  is the parameter that is used before optimization ranging from 0 – 1,  $Up_b$  and  $Lo_b$  are the upper and lower bounds respectively. Moreover, the optimization parameters  $K, H$  and  $B$  are computed by using the following equations:

$$K = U \times \left(1 - \frac{itr}{max_{itr}}\right) \quad (3)$$

$$U = \cos(2 \times \rho) + 1 \quad (4)$$

$$H = K \times a \quad (5)$$

where  $U$  is the function computed using equation (4), iteration denotes the current iteration, and  $\lceil \max \rceil$  iteration denotes the total number of iterations used to perform the optimization,  $\rho$  is the random number from 0 to 1, and  $a$  is the random number from -1 to 1. Consequently, the other optimization parameter  $B$  is estimated by using the following equation:

$$B = w \times G(h) \quad (6)$$

$$w = [-K, K] \quad (7)$$

Where,  $w$  indicates the random value in the range of [-K to k]. Similarly, the other two operators are used in the exploitation stage that supports to increase the performance of searching. During exploitation, the following operations are performed:

$$AG(h+1) = H \times Q \times (G(h) - G_{sb}) + G(h) \quad (8)$$

$$Q = \left( \left| \frac{1}{P} \sum_{i=1}^P AG_i(h) \right|^v \right)^{1/v} \quad (9)$$

$$v = 2^L \quad (10)$$

Where,  $G(h)$  indicates the vector position of gorilla,  $G_{sb}$  is the silverback of gorilla,  $Q$  is the function estimated by using equ (9),  $v$  is the setting parameter, and  $P$  is the number of gorillas. Then, the vector position of each candidate gorilla  $AG(i)$  is updated at iteration  $h$  as shown in below:

$$AG(i) = G_{sb} - (G_{sb} \times I - G(h) \times I) \times CV \quad (11)$$

$$I = 2 \times rand - 1 \quad (12)$$

Where,  $I$  is the impact force, and  $CV$  is the coefficient vector. Finally, the cost of all solutions is estimated, and based on this the best solution is obtained as the optimal solution.

---

*Algorithm 1 – GTO for feature selection*

---

Input: Population size  $P$ , maximum number of iterations  $N$ , parameters  $\delta$  and  $s$ ;  
Output: Location of gorilla and fitness value;  
Step 1: Initialize the random population as  $G_i$ , where  $i = 1, 2 \dots P$ ;  
Step 2: Estimate the fitness value of Gorilla;  
Step 3: While (Until reaching the stopping criterion) do  
    Update the parameter  $K$  by using equation (3);  
    Update the parameter  $H$  by using equation (5);  
    For (each Gorilla  $G_i$ ) do  
        Update the current location of gorilla by using equation (2);  
    End for;  
    Compute the fitness value;  
    If  $AG$  is better  $G$   
        Set  $G_{sb}$  as the best location;  
        For (each Gorilla  $G_i$ ) do  
            If ( $|K| \geq 1$ ) then  
                Update the current location of gorilla using equation (9);  
                Else  
                    Update the current location of gorilla by using equation (11);  
                End if;  
            End for;  
            Estimate the fitness value of gorilla;  
            If (New solution is greater than previous)  
                Update  $G_{sb}$  as the best position;  
            End while;  
Return  $G_{best}, Best_{fit}$ ;

---

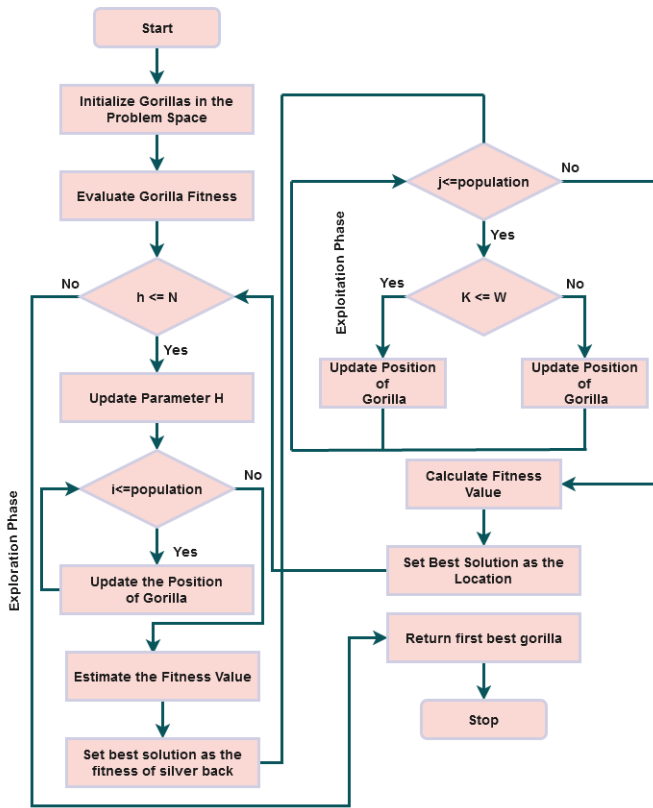


Fig. 5. Flow of GTOA.

#### E. Bi-Directional LSTM

Finally, the Bi-LSTM algorithm is applied to precisely classify the CVD with high reduced complexity and high accuracy. Typically, the Recurrent Neural Networks (RNNs) are the foundation of the LSTM method [29], Input gates, output gates, and forget gates that form the primary LSTM unit cell. The LSTM creates an internal feedback state due to processing various inputs close to one another, which helps the network comprehend time and the significant variability of the displayed data. Through gate control, the LSTM network incorporates resolving gradient vanishing to a certain level. While an LSTM framework can understand and preserve past data, it cannot accurately include new data to support a final prediction. As a result, a Bi-LSTM with two-way functionality was made using the LSTM network. In sequence prediction modelling, bi-LSTM networks perform better than other RNN and LSTM architectures, especially regarding speech or handwriting recognition and machine translation. Bi-LSTM contains two separate LSTMs that can integrate and aggregate data from both forward and backward directions. It is useful to have access to past and future context when performing sequence labeling tasks. Bi-LSTM proposes to transfer and invert each sequence to her two hidden entities, to produce a result by integrating the two logical states, and to obtain knowledge of the future and the past. Although they have identical sentence embedding, two opposing directions' parameters in the Bi-LSTM framework are different. By using this mechanism, the CVD is properly predicted and categorized with less time consumption.

#### IV. RESULTS AND DISCUSSIONS

In this section, we describe and compare the full results of the proposed disease prediction models. The CVD dataset is used to verify and contrast the outcomes of the suggested prediction mechanism. This data comprises extensive patient information and medical records. Moreover, the different types of parameters used to validate the results, and are computed as shown in below:

$$Precision = \frac{TrP}{TrP+FaP} \quad (13)$$

$$Recall = \frac{TrP}{TrP+FaN} \quad (14)$$

$$F1 - score = 2 \times \frac{(Precision \times Recall)}{(Precision+Recall)} \quad (15)$$

$$Accuracy = \frac{Correct Prediction}{Total Prediction} \quad (16)$$

Where,  $TrP$  is the true positive,  $FaP$  is the false positive, and  $FaN$  is the false negative. The attributes and descriptions of the CVD dataset used in this study are shown in Table I.

Fig. 6 presents a comparative analysis of conventional [30] and proposed AI mechanisms used for CVD prediction and classification. For this analysis, the parameters precision, recall, f-measure and accuracy have been considered. The proposed model offers improved computation capabilities and system efficacy as a result of the best feature processing and selection.

The accuracy of CVD prediction is significantly increased when the proposed optimization and classification model is used; however ordinary machine learning and deep learning methods perform less well due to insufficient feature processing and selection.

TABLE I. CVD DATASET DESCRIPTION

Attributes	Descriptions
Age	In years
Gender	Male or Female
Chest Pain	Type of CP
Blood Pressure	BP level in mm Hg
Serum Cholesterol	In mg/dl
Fasting blood sugar	>120 mg/dl (1 true and 0 false)
Rest ECG	0 – Normal 1 – Abnormal 2 – Maximum heart rate
Max_heart rate	Maximum heart rate
Exercise induced angina	0 – No 1 – Yes
ST depression	Depression induced by exercise
Slope	1- Up 2- Flat 3- Down
No of vessels	Vessels colored by fluoroscopy
Thalassemia	Normal, irreversible, and fixed
Num	No of risk, low risk, high risk, and very high risk

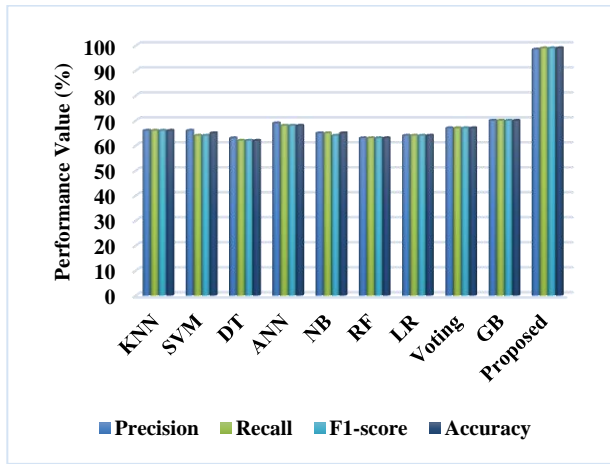


Fig. 6. Performance comparative analysis with machine learning techniques.

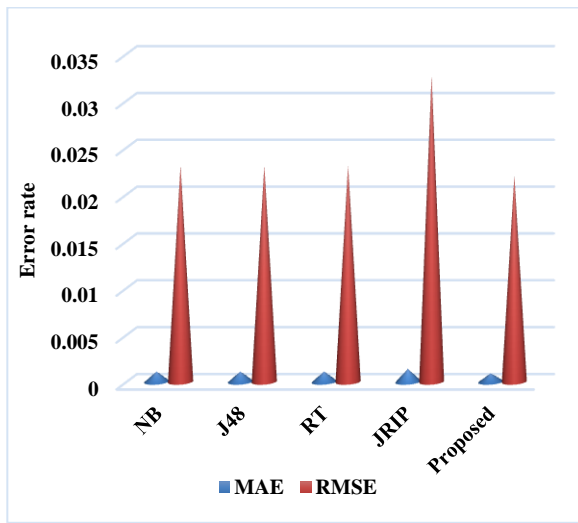


Fig. 7. Error analysis.

Fig. 7 compares the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) rate of the existing [31] and proposed classification methodologies. Absolute error is the term used to describe the size of a measurement error. Estimate using the difference between the measured value and the actual value. Mean Squared Error, or MSE, measures how inaccurate a statistical model is. The mean squared difference between observed and expected values is computed. Additionally, the MSE increases as model error increases and is equal to 0 when a model is error-free. Mean squared error, also called mean squared error, is a commonly used technique to assess the accuracy of predictions. It exemplifies the Euclidean gap between predictions and measured actual values, and is indicated in the results that the MAE.

Table II and Fig. 8 [32] provide an overall comparison of existing and proposed prediction methods used for CVD detection and classification in terms of recall, f1 score, accuracy, precision, and time shows the analysis. Based on our results, we conclude that the proposed BTOA-integrated Bi-LSTM model provides better results than other techniques. The GTOA helps to extract the most relevant and optimal features for classifying the disease. Hence, the training and

testing operations of the classifier are improved, so the overall prediction efficacy of the proposed model is highly improved.

Fig. 9 to Fig. 11 validates the CVD prediction performance of the baseline [33] and proposed mechanisms. Here, the similarity coefficients are also estimated for determining that how accurately the classifier predicts the disease from the dataset. From the overall results, it is determined that the combination of proposed GTOA integrated Bi-LSTM provides a highly improved results over the other techniques.

TABLE II. COMPARATIVE ANALYSIS

Methods	Recall	F1-score	Precision	Accuracy	Time (s)
DT	64.40	63.94	63.42	63.69	0.53
KNN	61.46	67.02	73.68	69.87	5.78
LR	67.99	71.13	74.58	72.36	2.52
NB	32.30	44.43	71.11	59.44	0.63
SVM	64.21	70.17	77.35	72.66	296.67
Proposed	99	99	99	99.5	0.32

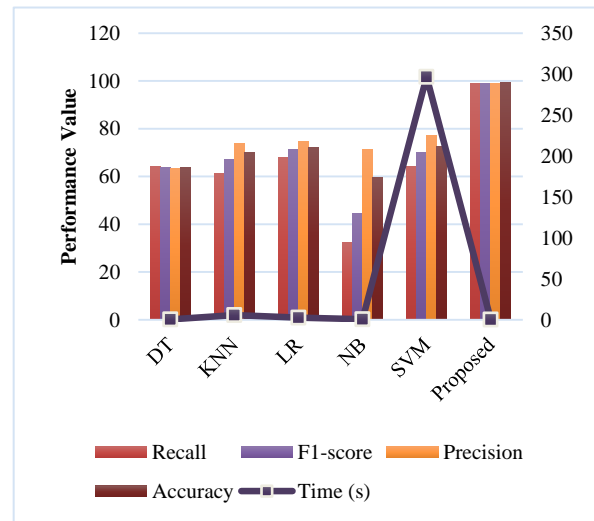


Fig. 8. Overall analysis.

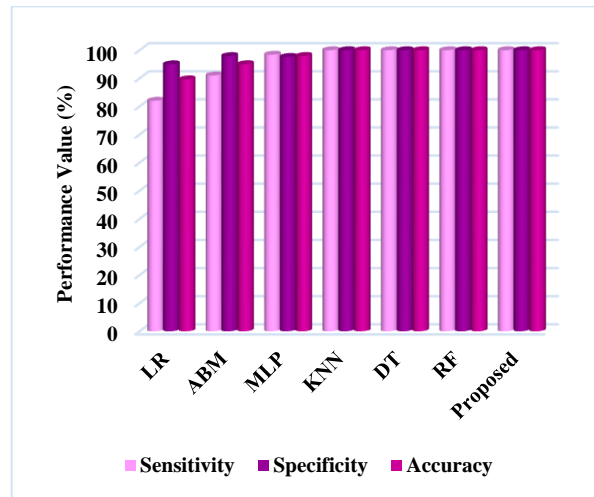


Fig. 9. Sensitivity, specificity, and accuracy analysis.



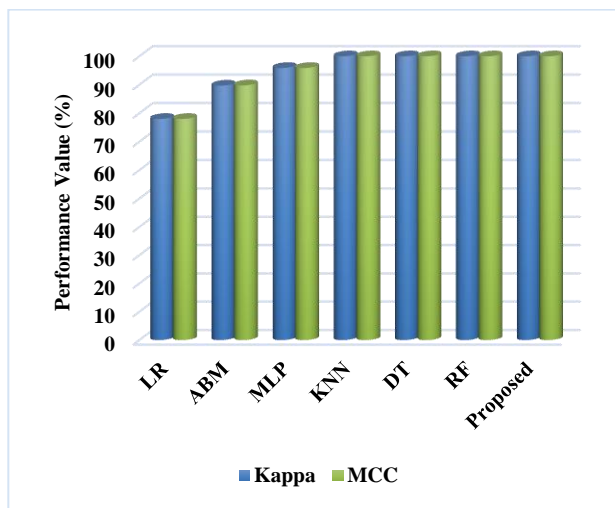


Fig. 10. Similarity coefficients.

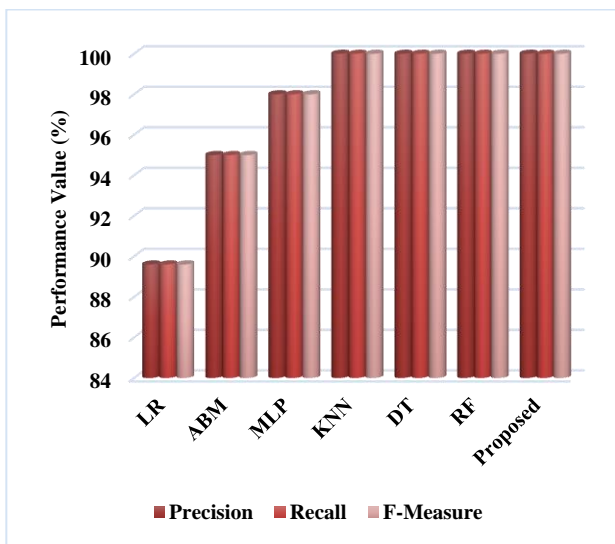


Fig. 11. Prediction performance evaluation.

## V. CONCLUSION

CVD is a life-threatening condition brought on by sudden heart problems, high blood pressure, and strokes. Early intervention and remote monitoring devices are the best ways to treat CVD patients. Also, the M-Health systems offer a variety of remote management applications for monitoring, and patients benefit from self-reported outcomes and affordable solutions. Recently, sophisticated and intelligent technologies can provide advanced procedures for healthcare services. Such technologies are also connected with m-Health systems to track, handle, and store patient data. This paper proposes a complete M-Health framework for predicting CVD using an advanced AI mechanism. The information collected by the sensors is sent to local databases and stored in cloud storage services with the latest technology enabling mobile networks. Data will be collected via a cloud computing platform or medical facility for further analysis. The GTOA-integrated Bi-LSTM techniques have been adapted for CVD classification. The suggested deep learning model is an excellent choice for CVD prediction based on examining

patient attributes from a dataset. During analysis, the performance and results of the proposed GTOA-integrated Bi-LSTM technique are validated and compared using various measures. Overall, the estimated results state that the proposed GTOA integrated with Bi-LSTM outperforms other methods with highly improved results.

## REFERENCES

- [1] O. S. Albahri, A. Zaidan, B. Zaidan, A. S. Albahri, A. H. Mohsin, K. Mohammed, et al., "New mHealth hospital selection framework supporting decentralised telemedicine architecture for outpatient cardiovascular disease-based integrated techniques: Haversine-GPS and AHP-VIKOR," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, pp. 219-239, 2022.
- [2] J. Divakaran, S. K. Prashanth, Gouse Baig Mohammad, Shitharth, Sachi Nandan Mohanty, C. Arvind, K. Srihari, Yasir Abdullah R., and Venkatesa Prabhu Sundramurthy, S. Shitharth, et al., "Improved handover authentication in fifth-generation communication networks using fuzzy evolutionary optimisation with nano core elements in mobile healthcare applications," in *Journal of Healthcare Engineering*, Hindawi, 2022, <https://doi.org/10.1155/2022/2500377>.
- [3] S. M. S. Islam and R. Maddison, "Digital health approaches for cardiovascular diseases prevention and management: lessons from preliminary studies," *Mhealth*, vol. 7, 2021.
- [4] N. Ji, T. Xiang, P. Bonato, N. H. Lovell, S.-Y. Ooi, D. A. Clifton, et al., "Recommendation to use wearable-based mhealth in closed-loop management of acute cardiovascular disease patients during the COVID-19 pandemic," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, pp. 903-908, 2021.
- [5] A. Joshi, I. Pant, and Y. Dhiman, "Efficient Analysis in Healthcare Domain using Machine Learning," in *Telemedicine: The Computer Transformation of Healthcare*, ed: Springer, 2022, pp. 125-134.
- [6] E. N. Schorr, A. D. Gepner, M. A. Dolansky, D. E. Forman, L. G. Park, K. S. Petersen, et al., "Harnessing mobile health technology for secondary cardiovascular disease prevention in older adults: a scientific statement from the American Heart Association," *Circulation: Cardiovascular Quality and Outcomes*, vol. 14, p. e000103, 2021.
- [7] J. Calvillo-Arbizu, D. Naranjo-Hernández, G. Barbarov-Rostán, A. Talamino-Barroso, L. M. Roa-Romero, and J. Reina-Tosina, "A Sensor-Based mHealth Platform for Remote Monitoring and Intervention of Frailty Patients at Home," *International journal of environmental research and public health*, vol. 18, p. 11730, 2021.
- [8] S.-H. Kang, H. Baek, J. Cho, S. Kim, H. Hwang, W. Lee, et al., "Management of cardiovascular disease using an mHealth tool: a randomized clinical trial," *NPJ digital medicine*, vol. 4, pp. 1-7, 2021.
- [9] E. Epstein, N. Patel, K. Maysent, and P. R. Taub, "Cardiac rehab in the COVID era and beyond: mHealth and other novel opportunities," *Current cardiology reports*, vol. 23, pp. 1-8, 2021.
- [10] Hana Almagrabi, Abdulrhman M. Alshareef, Hariprasath Manoharan et al., "Empirical Compression Features of Mobile Computing and Data Applications Using Deep Neural Networks," *Security and Communication Networks*, vol. 2022, Article ID 8125494, 11 pages, 2022, <https://doi.org/10.1155/2022/8125494>.
- [11] L. Zhu, N. Li, L. Sun, D. Zheng, and G. Shao, "Non-coding RNAs: The key detectors and regulators in cardiovascular disease," *Genomics*, vol. 113, pp. 1233-1246, 2021.
- [12] A. H. Khan, M. Hussain, and M. K. Malik, "Cardiac disorder classification by electrocardiogram sensing using deep neural network," *Complexity*, vol. 2021, 2021.
- [13] M. Padmaja, S. Shitharth, K. Prasuna, A. Chaturvedi, P. R. Kshirsagar, and A. Vani, "Grow of artificial intelligence to challenge security in IoT application," *Wireless Personal Communications*, pp. 1-17, 2021.
- [14] R. S. Istepanian and T. Al-Anzi, "m-Health 2.0: new perspectives on mobile health, machine learning and big data analytics," *Methods*, vol. 151, pp. 34-40, 2018.
- [15] K. N. Mishra and C. Chakraborty, "A novel approach towards using big data and IoT for improving the efficiency of m-health systems," in

- Advanced computational intelligence techniques for virtual reality in healthcare, ed: Springer, 2020, pp. 123-139.
- [16] S. R. Alotaibi, "Applications of artificial intelligence and big data analytics in m-health: a healthcare system perspective," *Journal of healthcare engineering*, vol. 2020, 2020.
- [17] A. Al-Marridi, A. Mohamed, and A. Erbad, "Ai-based techniques on edge devices to optimize energy efficiency in m-health applications," in *Energy Efficiency of Medical Devices and Healthcare Applications*, ed: Elsevier, 2020, pp. 1-23.
- [18] S. Elhishi, S. Alzaky, A. El-Metwally, B. Burham, S. Ragab, S. Elgayar, et al., "Leu-Life: A Smart Application for Leukemia Cancer Patients Based on Machine Learning," 2022.
- [19] K. Lano, S. Y. Tehrani, M. Umar, and L. Alwakeel, "Using Artificial Intelligence for the Specification of m-Health and e-Health Systems," in *The Future Circle of Healthcare*, ed: Springer, 2022, pp. 273-299.
- [20] I. R. Mendo, G. Marques, I. de la Torre Díez, M. López-Coronado, and F. Martín-Rodríguez, "Machine learning in medical emergencies: a systematic review and analysis," *Journal of Medical Systems*, vol. 45, pp. 1-16, 2021.
- [21] C. Pankaj, K. V. Singh, and K. R. Singh, "Artificial Intelligence enabled Web-Based Prediction of Diabetes using Machine Learning Approach," in *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*, 2021, pp. 60-64.
- [22] A. S. Abed, B. Khalil, S. Ibrahim, M. A. Zahra, M. A. Salih, and R. A. Jaleel, "Development of an Integrate E-Medical System Using Software Defined Networking and Machine Learning," *Webology*, vol. 19, pp. 3410-3418, 2022.
- [23] M. Alhussein and G. Muhammad, "Voice pathology detection using deep learning on mobile healthcare framework," *IEEE Access*, vol. 6, pp. 41034-41041, 2018.
- [24] A. Shaban-Nejad, M. Michalowski, and D. L. Buckeridge, "Health intelligence: how artificial intelligence transforms population and personalized health," vol. 1, ed: Nature Publishing Group, 2018, pp. 1-2.
- [25] A. B. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: a scoping review of methods and applications," *Psychological medicine*, vol. 49, pp. 1426-1448, 2019.
- [26] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A survey of deep learning and its applications: a new paradigm to machine learning," *Archives of Computational Methods in Engineering*, vol. 27, pp. 1071-1092, 2020.
- [27] E. Garcia-Ceja, M. Riegler, T. Nordgreen, P. Jakobsen, K. J. Oedegaard, and J. Tørresen, "Mental health monitoring with multimodal sensing and machine learning: A survey," *Pervasive and Mobile Computing*, vol. 51, pp. 1-26, 2018/12/01/ 2018.
- [28] S. Tian, W. Yang, J. M. Le Grange, P. Wang, W. Huang, and Z. Ye, "Smart healthcare: making medical care more intelligent," *Global Health Journal*, vol. 3, pp. 62-65, 2019.
- [29] S. Banik, N. Sharma, M. Mangla, S. N. Mohanty, and S. Shitharth, "LSTM based decision support system for swing trading in stock market," *Knowledge-Based Systems*, vol. 239, p. 107994, 2022.
- [30] H. H. Alalawi and S. A. Manal, "Detection of Cardiovascular Disease using Machine Learning Classification Models," *International Journal of Engineering Research & Technology (IJERT)* ISSN, pp. 2278-0181, 2021.
- [31] R. G. Nadakinamani, A. Reyana, S. Kautish, A. S. Vibith, Y. Gupta, S. F. Abdelwahab, et al., "Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques," *Computational Intelligence and Neuroscience*, vol. 2022, p. 2973324, 2022/01/11 2022.
- [32] W. M. Jinjri, P. Keikhosrokiani, and N. L. Abdullah, "Machine Learning Algorithms for The Classification of Cardiovascular Disease-A Comparative Study," in *2021 International Conference on Information Technology (ICIT)*, 2021, pp. 132-138.
- [33] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Computers in Biology and Medicine*, vol. 136, p. 104672, 2021.

# Reducing Cheating in Online Exams Through the Proctor Test Model

The Case of Indonesian Learners of Arabic from the Book of Silsilat Al-Lisan

Yusring Sanusi Baso<sup>1\*</sup>, Nurul Murtadho<sup>2</sup>, Syihabuddin<sup>3</sup>, Hikmah Maulani<sup>4</sup>  
Andi Agussalim<sup>5</sup>, Haeruddin<sup>6</sup>, Ahmad Fadlan<sup>7</sup>, Ilham Ramadhan<sup>8</sup>

Department of Arabic Studies-Faculty of Cultural Science, Hasanuddin University, Makassar, Indonesia 90245<sup>1</sup>

Department of Arabic Studies, Malang State University, Indonesia<sup>2</sup>

Department of Arabic Studies, Indonesia University of Education, Indonesia<sup>3,4</sup>

Department of Arabic Studies-Faculty of Cultural Science, Hasanuddin University, Makassar, Indonesia 90245<sup>5,6,7,8</sup>

**Abstract**—The World Health Organization (WHO) officially declared coronavirus (COVID-19) a pandemic on March 11, 2020. Educational institutions must change most face-to-face learning activities in class to online. This situation forces academic institutions to change the format of assessing student learning outcomes. Online exam surveillance applications utilizing cameras and other blocking browsers (proctors) are becoming popular. However, the appearance of the proctor model supervised exam system also raises controversy. The main discussion regarding this proctor system is the integrity of assessment and the capacity of students to adapt to this new method of supervision. The main question is whether students feel comfortable using the proctor system in exams and whether this system affects students' scores. To answer this question, we have analyzed the scores obtained from a trial of 152 scores of students learning Arabic at Hasanuddin University Makassar, Indonesia. The experiment involved three exam models: online format from home using the Sikola Learning Management System (Modality 1), online directly using the Proctor System in the Sikola Learning Management System (Modality 2), and a paper exam format in person under the supervision of a lecturer (Modality 3). The results show that students prefer Modality 1 (online at home with the Sikola LMS system). There is a statistical difference between the scores obtained by students from the three modalities analyzed. Student scores with modality 1 are higher than the other two modalities. On the other hand, there was no difference in scores between modalities 2 and 3. The online exam system (modality 2) can be applied to online exams in higher education institutions because it can reduce or even keep students from cheating.

**Keywords**—Reducing cheating; online exam; proctor test models; Indonesian learners of Arabic; Silsilat Al-Lisan

## I. INTRODUCTION

The World Health Organization (WHO) declared the new coronavirus (COVID-19) a pandemic on March 11, 2020 [1] officially. COVID-19 has spread rapidly due to the high transmission capacity of the virus and the routes of transmission (especially via aerosols when coughing and sneezing). In addition, about 30% of patients have various life-threatening symptoms [2]. Higher education institutions must change their learning process activities based on the spread and the symptoms it causes. The face-to-face learning process is suspended. Universities are trying to adapt to this

situation which has become a challenge worldwide. At the study program level, the teaching team must adjust their teaching and assessment systems [3]. On the other hand, lecturers and students have received online learning activities well, both in learning activities and in the assessment system.

An online assessment or exam system with strict monitoring is starting to be widely used. This proctor system has generated significant controversy, especially in science education [4]–[7]. Among the concerns of educators and students are psychological disturbances, privacy, and various environmental factors [5]. Dragan et al. (2020) concluded that online exams with remote supervision proved to be a timely solution, and also, the emotional needs of students who might feel stressed by these rigorous proctor exams should be considered [8].

Two of the most significant challenges in the online exam proctor system are the assessment standards and students' ability to adapt to this new exam modality. Several studies have been carried out to analyze student dishonesty in the online exam proctor system and various ways to prevent it. A study by Guangul et al. (2020) concluded that combining various assessment methods has helped minimize academic dishonesty [4]. Li (2021) developed an anti-collusion approach based on optimizing remote online testing [9]. Recently, Pettit et al. (2021) have analyzed studies conducted so far that provide recommendations for improving student authentication and preventing fraud [7]. Baso [2022] also wrote an online exam model with a reliable proctor system to reduce academic cheating that might occur [10].

The integrity of the evaluation in online exams can be monitored in various ways, including taking the exam in person or using a real-time supervisor system [5]. One of the real-time online monitoring systems is the website Uji.sikoola.com which provides real-time online monitoring services using a webcam and a browser lock [11]. Students are connected to the online exam page sikoola.ujian.com which monitors students during the online exam. Before the exam starts, students must check their internet speed. The sikoola.ujian.com system will analyze the internet speed used by students in real-time and provide good, moderate, or bad status for student internet speed. Furthermore, students must

\*Corresponding Author.

check the webcam device they are using. If the webcam is not working, students cannot enter the online exam waiting room.

The main objective of this study was to compare the attitudes and scores obtained by students in the Arabic study program who took the Arabic language skills exam learned from the *Silsilat al-Lisan* book [12] through the online exam proctor system on the *sikoola.ujian.com* page.

In relation to the purpose of this research, the questions to be answered are:

- What is the students' attitude towards the three modalities of the exam in measuring the Arabic language skills they have learned from the book *Silsilat al-Lisan*?
- Are there differences in the scores obtained by students from the three exam modalities given?

## II. MATERIALS AND METHODS

### A. Population and Sample

The research was conducted in the Arabic study program at Hasanuddin University, in two batches of students, namely the 2022 and 2021 batches (ages between 20-21 years). The 2022 batch consists of 52 students, and the 2021 batch of 46 students. This group of students sequentially took Arabic 1 and Arabic 3 courses. The questions tested were sourced from the book *Silsilat al-Lisan*.

The exam is carried out three times at an interval of two weeks. The score range of each test is 0-100. The model of the three exam questions is the Multiple-Choice Question (MCQ). The first exam was conducted online via <https://sikola.unhas.ac.id>. Students take exams simultaneously where they can access this page from anywhere (modality 1). The second exam is also conducted online on the <https://sikoola.ujian.com> page which has a strict monitoring feature (modality 2). The third exam was conducted in class with supervision from a team of lecturers supporting Arabic 1 and Arabic 2 courses (modality 3).

### B. Study Design

Students must take all three exams (modalities 1, 2, and 3). The reason is that students can feel the difference when taking exams with different modalities. In the modality one exam, they take the exam from anywhere. It's just that the time for carrying out the exam has been determined when it starts to be accessed and when the questions will be closed automatically. In other words, the first modality test is carried out simultaneously at the same time.

Examination with modality two is also carried out simultaneously at the same time. Exam venues may vary. It's just that there are mandatory requirements that must be met, namely, the device used must have a functioning webcam. In addition, the internet speed used is at least 40 Mbps upload. This speed would be considered moderate. If the internet speed test result equals or exceeds 50 Mbps, the system will categorize it as good. However, if the internet speed test results are below 40 Mbps, the *sikoola.ujian.com* system will assess the network used by this student as bad. Bad speed test results will prevent students from accessing exam questions on

the *sikoola.ujian.com* page. If this condition occurs, students must move locations to find an access point with a minimum internet speed test of 40 Mbps.

In addition, in this online exam with modality 2, the exam supervisor can observe the behavior of the examinees. The faces of each examinee will be displayed on the monitor screen on the *sikoola.ujian.com* page. If there are examinees who move a lot by looking left and right, the exam supervisor can send a warning message to these students. In addition, if an examinee opens a new page other than the exam page from the browser used, the *sikoola.ujian.com* application will block it. Thus, the examinee's monitor screen will be locked. Likewise, if the examinee opens another browser, the screen will be locked again. In the same way, if the examinee presses a key on the keyboard, the monitor screen will also be locked. Lockdown time can be set as needed, for example, 20 seconds or 50 seconds, or even an hour.

Examination with modality three is conducted in class. Examinees receive test papers. The exam supervisor will supervise them during the test. After students took the three MCQ tests with three modalities, they filled out a survey that included questions about their attitudes toward the three test modalities.

### C. Statistical Analysis

The research team used the statistical software SPSS 26 to analyze the data. The research team conducted a normality test on standardized residual values to calculate whether the three test score variables were normal or not. If the results are normal, then the repeated measure ANOVA test is used, and vice versa. If it is not normal, the researcher will use the Friedman (non-parametric statistics) test to analyze the data.

The basis for the decision of the normality test used in this study is:

- If the Sig value  $> 0.05$ , then the data is normally distributed.
- If the Sig value  $< 0.05$ , the data are not normally distributed.

The interpretation is as follows:

- The Within-Subjects Factors output table shows three score variables from three exams (modality 1, modality 2, and modality 3).
- The output table of Mauchly's Test of Sphericity is used to see the similarity of the assumption of variance (Sphericity Assumed) of the research data provided that the Sig value is  $< 0.05$ . If these conditions are not met, the researcher will use the Greenhouse-Geisser as a condition for testing the hypothesis.
- The Tests of Within-Subjects Effects output table will be used by researchers to:

- 1) Make a hypothesis formulation.
- 2) Know the basis for decision making.
- 3) Conclusion.

- The formulation of the hypothesis is:
  - H0: there is no difference in the average score of the three test modalities given to students
  - Ha: there is a difference in the average score of the three test modalities given to students.
  - The basis for the decision to be used is:

- 1) If the value of Greenhouse-Geisser Sig > 0.05, then H0 is accepted and Ha is rejected.
- 2) If the Greenhouse-Geisser Sig value < 0.05, then H0 is rejected and Ha is accepted.

### III. RESULTS

A descriptive analysis of this study can be seen in Table I:

TABLE I. DESCRIPTIVE

			Statistic	Std. Error
Standardized Residual for MODALITY1	Mean		0,0000	0,10102
	95% Confidence Interval for Mean	Lower Bound	-0,2005	
		Upper Bound	0,2005	
	5% Trimmed Mean		-0,0022	
	Median		-0,2304	
	Variance		1,000	
	Std. Deviation		1,00000	
	Minimum		-1,61	
	Maximum		1,70	
	Range		3,30	
	Interquartile Range		1,65	
	Skewness		-0,019	0,244
	Kurtosis		-1,291	0,483
	Standardized Residual for MODALITY2	Mean		0,0000
95% Confidence Interval for Mean		Lower Bound	-0,2005	
		Upper Bound	0,2005	
5% Trimmed Mean		-0,0007		
Median		0,0324		
Variance		1,000		
Std. Deviation		1,00000		
Minimum		-1,69		
Maximum		1,75		
Range		3,44		
Interquartile Range		1,72		
Skewness		-0,034	0,244	
Kurtosis		-1,186	0,483	
Standardized Residual for MODALITY3		Mean		0,0000
	95% Confidence Interval for Mean	Lower Bound	-0,2005	
		Upper Bound	0,2005	

5% Trimmed Mean	-0,0058	
Median	-0,1034	
Variance	1,000	
Std. Deviation	1,00000	
Minimum	-1,53	
Maximum	1,68	
Range	3,22	
Interquartile Range	1,67	
Skewness	0,196	0,244
Kurtosis	-1,178	0,483

#### A. Research Question One

The data is obtained in the following Table II to answer the first question of this study about students' attitudes towards the results of the MCQ exam scores with three modalities.

TABLE II. ATTITUDES TOWARDS MODALITIES

NUMBER OF STUDENTS	ATTITUDE TOWARDS MODALITIES					
	MODALITY 1		MODALITY 2		MODALITY 3	
	Trust	Untrust	Trust	Untrust	Trust	Untrust
Batch 2021	16	30	46	0	46	0
Batch 2022	33	35	52	0	52	0

#### B. Research Question Two

Table III shows three variables to be tested, namely, MCQ score modality 1, MCQ score modality 2, and MCQ score modality 3, as shown in Table IV.

As for Sphericity, Assumed with the condition that Sig. < 0.05 and then in the following Table V, it is obtained Sig 0.010 > 0.05. Thus, Sphericity Assumed is not fulfilled. Therefore, Greenhouse-Greiser will be used as the basis for testing the hypothesis.

TABLE III. TESTS OF NORMALITY

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual for MODALITY1	0,173	98	0,000	0,934	98	0,000
Standardized Residual for MODALITY2	0,105	98	0,010	0,951	98	0,001
Standardized Residual for MODALITY3	0,092	98	0,039	0,937	98	0,000

a. Lilliefors Significance Correction

TABLE IV. WITHIN-SUBJECTS FACTORS

Measure:	MCQSCORE
MODALITY	Dependent Variable
1	MODALITY1
2	MODALITY2
3	MODALITY3

TABLE V. MAUCHLY'S TEST OF SPHERICITY

Mauchly's Test of Sphericity <sup>a</sup>							
Measure:	MCQSCORE						
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon <sup>b</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
MODALITY	0,908	9,283	2	0,010	0,916	0,932	0,500
Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.							
a. Design: Intercept Within Subjects Design: MODALITY							

To get conclusions from this statistical data, it can be seen in Table VI. This output table, known as Tests of Within-Subjects Effects, will display the Greenhouse-Geisser as can be seen in Table V below:

TABLE VI. TESTS OF WITHIN-SUBJECTS EFFECTS

Measure:	MCQSCORE					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	
MODALITY	Sphericity Assumed	17880,660	2	8940,330	193,272	0,000
	Greenhouse-Geisser	17880,660	1,831	9764,337	193,272	0,000
	Huynh-Feldt	17880,660	1,865	9589,111	193,272	0,000
	Lower-bound	17880,660	1,000	17880,660	193,272	0,000
Error(MODALITY)	Sphericity Assumed	8974,007	194	46,258		
	Greenhouse-Geisser	8974,007	177,628	50,521		
	Huynh-Feldt	8974,007	180,874	49,615		
	Lower-bound	8974,007	97,000	92,516		

Based on the hypothesis with the Greenhouse-Geisser Sig. < 0.05, then 0.000 < 0.05 is obtained, meaning H0 is rejected, and Ha is accepted, or there is a difference in the average score of the three test modalities given to students.

IV. DISCUSSION

The COVID-19 pandemic has forced lecturers and students to change academic activities, including activities to measure student learning outcomes. Several research results show acceptance of the learning achievement measurement method with proctor applications [13]–[15], including in Indonesia [11]. In this study, to answer the first research question, the results obtained can be seen in Table I previously, where it can be seen that students believe in the MCQ exam model modalities 2 and 3. However, in the 2021 and 2022 batches, some students do not believe in modalities 1.

The research team conducted interviews to explore the causes of the distrust of some students towards the MCQ exam

model with modality 1. In general, it can be concluded that the cause of this distrust is the opportunity for students to commit fraud during the exam. Opportunities for students to commit fraud include searching for answers via Google (Googling), asking other friends via mobile (chatting via WhatsApp, Telegram, etc.) during exams, or opening books. Opportunities for fraud are open because there is no direct supervision. However, some students still have a reliable academic attitude by not cheating during exams.

On the other hand, the research team found that the basis for the 100% confidence of the two groups of students in the MCQ modality 2 and 3 exam models was the feelings of students monitored both offline and online. This strict supervision makes students not think of committing fraud. This condition instead makes students focus on answering exam questions.

Table I above (descriptive statistics) shows that the test scores in modality 1 are higher than in modalities 2 and 3. This data can be accepted logically because it is possible that during the exam, some students cheated. They are looking for answers to questions they have difficulty answering via googling or asking other friends by chatting. On the other hand, the test scores on modalities 2 and 3 are lower because, logically, it can also be understood that there are no opportunities for students to commit fraud. Examination modalities 2 and 3 close opportunities for cheating during exams, especially for students who have cheated on exams with modality 1.

The assumptions of the research team are based on Fig. 1:

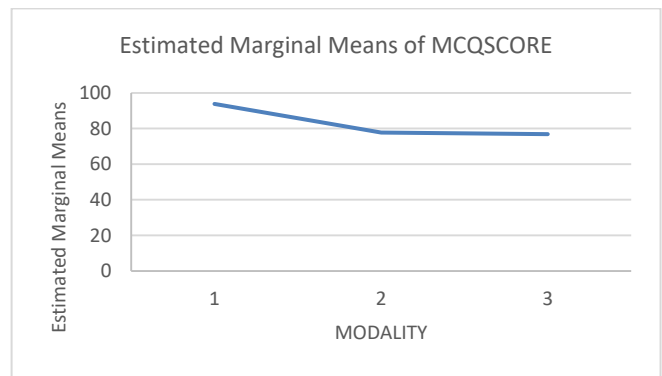


Fig. 1. The test scores with modality.

Fig. 1 shows that the results of the test scores with modality 1 are higher than modalities 2 and 3. However, the results of the test scores for modalities 2 and 3 do not appear to have a significant difference. On the other hand, modality 2 and modality 3 have different test channels; modality 2 is carried out online while modality 3 is carried out offline.

Logically the score results are relatively the same or not significantly different even though the exam channels are different because exam students are under strict supervision. In modality 2, supervision uses the proctor application system. As for modality 3, supervision is carried out directly by F2C in the exam room.

Thus it can be said that modalities 2 and 3 have in common close supervision during the exam. This strict supervision puts the exam atmosphere in a safe condition where students think they have no opportunity to cheat in the exam.

## V. CONCLUSIONS

In this study, it was found that the attitude of students lacked confidence in the score of the exam results with modality 1. The cause of this distrust was caused by the opportunity for students to commit fraud during the exam. In contrast, students believe in the score of the exam results with modalities 2 and 3. Student trust in the score of the exam results with modalities 2 and 3 is due to the strict supervision conditions during the exam even though the channels of modality 2 and 3 are different, where modality 2 is online channel whereas modality 3 is offline channel.

With these results, it can be claimed that measuring learning outcomes through MCQ with modalities 2 and 3 is reliable. If the exam is conducted online, modality 2 can be used as a solution; if offline, modality 3 is used as an alternative.

## ACKNOWLEDGMENTS

We are grateful to the all students of Arabic Study program of Universitas Hasanuddin who involved in this research.

## AUTHOR CONTRIBUTIONS

Conceptualization, Yusring Sanusi Baso; methodology, Yusring Sanusi Baso, Hikmah Maulani, dan Andi Agussalim; formal analysis, Yusring Sanusi Baso and Haeruddin; investigation, Fadlan Ahmad; data curation, Ilham Ramadhan; writing-original draft preparation, Yusring Sanusi Baso; writing-review and editing, Syihabuddin anda Nurul Murtadho. All authors have read and agreed to the published version of the manuscript.

## REFERENCES

- [1] Cucinotta and M. Vanelli, "WHO declares COVID-19 a pandemic," *Acta Biomed.*, vol. 91, no. 1, pp. 157–160, 2020.

- [2] P. J. Marín García, A. Arnau, and L. Llobat, "Preferences and scores of different types of exams during COVID-19 pandemic in faculty of veterinary medicine in Spain: A cross-sectional study of paper and E-exams," *Educ. Sci.*, vol. 11, no. 8, pp. 0–5, 2021.
- A. S. Milone, A. M. Cortese, R. L. Balestrieri, and A. L. Pittenger, "The impact of proctored online exams on the educational experience," *Curr. Pharm. Teach. Learn.*, vol. 9, no. 1, pp. 108–114, 2017.
- [3] F. M. Guangul, A. H. Suhail, M. I. Khalit, and B. A. Khidhir, "Challenges of remote assessment in higher education in the context of COVID-19: a case study of Middle East College," *Educ. Assessment, Eval. Account.*, vol. 32, no. 4, pp. 519–535, 2020.
- [4] F. F. Kharbat and A. S. Abu Daabes, "E-proctored exams during the COVID-19 pandemic: A close understanding," *Educ. Inf. Technol.*, vol. 26, no. 6, pp. 6589–6605, 2021.
- [5] P. R. Lockman, J. A. Gaasch, K. Borges, A. Ehlo, and Q. R. Smith, "Using WebCt to implement a basic science competency education course," *Am. J. Pharm. Educ.*, vol. 72, no. 2, p. 39, 2008.
- [6] M. Pettit, S. Shukla, J. Zhang, K. H. Sunil Kumar, and V. Khanduja, "Virtual exams: has COVID-19 provided the impetus to change assessment methods in medicine?," *Bone Jt. Open*, vol. 2, no. 2, pp. 111–118, 2021.
- [7] I. F. Dragan, L. F. Yildiz, K. Dunn, and A. Ramesh, "Integrating remote proctoring in dental education: Problem, solution, and results," *J. Dent. Educ.*, vol. 85, no. S1, pp. 1071–1073, 2021.
- [8] M. Li, "Optimized collusion prevention for online exams during social distancing," *npj Sci. Learn.*, vol. 6, no. 1, 2021.
- [9] S. A., "Role of Zinc Nutrition for Increasing Zinc Availability, Uptake, Yield, and Quality of Maize (*Zea Mays* L.) Grains: An Overview," *Communications in Soil Science and Plant Analysis*, vol. 51, no. 15, pp. 2001–2021, 2020.
- [10] Y. S. Baso, "Proctoring and Non-proctoring Systems: A Comparative Study of Online Exam Scores for an Arabic Translating Course," vol. 13, no. 6, pp. 75–82, 2022.
- [11] D. M. A. and M. B. Dr. Bakri Sheigh Ameen, Ali Hamadallah, Dr. Fakhruddin Qabawah, Dr. Mazen Almubarak, "Kitab Silsilat Al-Lisan," Abu Dhabi, 2021. [Online]. Available: <https://mothertongue.ae/arabic-curricula/silsilat-al-lisan/>. [Accessed: 06-Jun-2022].
- M. S. Medina and A. N. Castleberry, "Proctoring strategies for computer-based and paperbased tests," *Am. J. Heal. Pharm.*, vol. 73, no. 5, pp. 274–277, 2016.
- [12] H. Meishar-Tal and A. Levenberg, "In times of trouble: Higher education lecturers' emotional reaction to online instruction during COVID-19 outbreak," *Educ. Inf. Technol.*, vol. 26, no. 6, pp. 7145–7161, 2021.
- [13] A. Sangwan, A. Sangwan, and P. Punia, "Development and Validation of an Attitude Scale towards Online Teaching and Learning for Higher Education Teachers," *TechTrends*, vol. 65, no. 2, pp. 187–195, 2021.

# A Framework of Outcome-based Assessment and Evaluation for Computing Programs

Wasan S. Awad, Khadija A. Almhoson  
College of Information Technology, Ahlia University, Manama, Bahrain

**Abstract**—This paper is to present a framework for student outcome-based assessment and evaluation, including the process and detailed activities leading to continue assessment of the successes of an academic program which is essential to its sustainability. Moreover, this paper provides a survey of the literature that reviews the different means of assessing and evaluating an academic program together with the critical performance metrics which aid in quantifying such evaluation. The presented framework is implemented on the Information Technology program over a course of five years. The paper provides empirical insights about how careful implementation of the presented framework enabled the College of Information Technology in Ahlia University to achieve outstanding results in quality assurance and to be ABET accredited. The results of the implementation prove the effectiveness of the framework in improving the student performance and the program. This paper fulfils an identified need to study how student outcome-based assessment and evaluation model enables an academic institute to foster quality assurance instead of relying on ad hoc practices which might lead them to trial-and-error approach. The presented framework could be followed by other institution aiming for international accreditations.

**Keywords**—Student outcomes; program assessment; program evaluation; program accreditation; ABET accreditation; continuous improvement

## I. INTRODUCTION

The College of Information Technology (IT) of Ahlia University has a clear academic planning framework to fulfill the aims and objectives of its academic programs. The academic planning framework is transformed into an annual college operational plan that is derived from the strategic plan of the University. The academic planning framework for Information Technology Program comprises three perspectives. One is the study plan that has to be carried out in order to execute the program successfully to the students enrolled in the program. The second perspective is the Program review. The purpose of the quality program review plan is to apply measures that ensure the quality of delivery, assessments, and contents of the program. Third is the college operational plan that is aligned with university strategic plan and is meant to conduct activities/tasks that provide continuous support to the successful and smooth execution of the program.

The Information Technology (BSIT) program, as any other academic program, leading to an academic degree, is based on a curriculum constituting of a combination of courses and relevant activities that are organized for the achievement of some learning outcomes defined by the offering institute/university. In this regard, the program shall have well defined

Program Educational Objectives (PEOs) that leads to fulfilling the mission and goals of both college and university. These objectives should:

- be broad statements which express what graduates are expected to attain after few years of graduation [1], [2];
- serve the missions of both college and university;
- satisfy program constituencies needs.

Furthermore, the set of Student Outcomes (SOs) of the program should be defined. These SOs describe what students are expected to know and be able to do by the time of graduation. These relate to the skills, knowledge and behaviors that students acquire as they progress through the program. There are six SOs defined by Accreditation Board for Engineering and Technology (ABET) [1].

At the level of courses of BSIT, students' performance is measured by the achievement of relevant course level outcomes. These course level outcomes are called Course Intended Learning Outcomes (CILOs). The CILOs are measured against the performance indicators set at IT program level. These performance indicators are generated from the student outcomes defined for the program. CILOs are satisfying the standards set by Bahrain National Qualification Framework (NQF) of the level 8. The courses in BSIT program have up to thirteen (13) CILOs that are required to be achieved by the students in each individual course. The achievements of the CILOs in all the courses translate the achievement of student outcomes for the program. These CILOs are related to the Knowledge and Understanding (A1, A2, A3), Subject-Specific Skills (B1, B2, B3), Critical-Thinking Skills (C1, C2, C3), and General and Transferable Skills (D1, D2, D3, D4) that students acquire as they progress through the program.

The program assessment and evaluation process should take place at different levels, including PEOs, SOs and CILOs. This ensures having a comprehensive and effective process in place. Regarding program assessment, it is the practice of identifying, collecting, and preparing the data needed for the evaluation. Regarding evaluation, it could be viewed as the practice of interpreting the data collected through the different assessment means to find the attainment of the SOs [3].

The objective of this paper is to thoroughly present and document a framework based on SOs assessment and evaluation that could be adopted by any academic program for the purpose of continuously assessing its success level. Thus, this paper presents the well-defined procedure and detailed activities comprising the framework developed by the College



of IT of Ahlia University for assessing and evaluating student outcomes of BSIT program- that are derived from ABET SOs- to fulfil the college mission and goals, achieve program objectives as well as achieve the prescribed graduate attributes by attaining the student outcomes. This framework is based on the longitudinal methodology followed while applying SO based assessment and evaluation model on BSIT program over a course of five years. It is intended to remain valid even if ABET redefined it's SOs. The findings will guide other higher education institutions to continuously evaluate and improve their academic programs, achieve outstanding results in quality assurance, and to be ABET accredited.

The remainder of the paper is organized as follows. Section II presents a detailed survey of the literature highlighting different means of assessing and evaluating academic programs together with the performance metrics in use. Section III provides a comprehensive overview of the proposed framework followed by the College of IT in Ahlia University. Section IV discusses the results of implementing the framework over five academic semesters. Section V lists the challenges faced by the college while implementing the framework and the college actions in overcoming these challenges. Last section presents the conclusion.

## II. LITERATURE REVIEW

### A. Survey of the Literature

Literature is rich with studies on the different approaches of program assessment and evaluation. One of the common approaches is curriculum mapping. It explores the links between curriculum content to be delivered and the learning outcomes. This is done by inspecting the relationship between the intended, delivered and received curriculum. The intended curriculum could be examined through the courses' syllabi and materials prepared by faculty members, which contain the learning objectives for each course. The taught curriculum could be examined through interviews with faculty members teaching/ coordinating the courses. The received curriculum could be explored through focus group interviews with students. This approach suggests to use student learning outcome as the base of curriculum development and assessment by treating it as the basis for formative and summative assessments. It considers when, how, and what is taught, as well as the assessment measures in use to gauge achievement of expected student learning outcomes. The strength of this approach is that it aids in identifying whether the intended curriculum and courses' materials has been actually taught and has been actually learnt by students. Moreover, it demonstrates the links between the main components of the curriculum, which are: learning outcomes, learning opportunities, content, and assessment. Furthermore, it studies the curriculum from multiple perspective. On the other hand, this approach has some drawbacks. The main one is being subject to recall bias from faculty members and students. Added to that, it is impossible to determine the extent to which topics were reinforced in the curriculum [4].

Another approach [5] is presented for the ABET-accredited program. Their strategy is based on the assessment and evaluation of their program at both course and program levels. Their process started by defining the vision and mission of

their program that serves the university vision. Then they established PEOs that leads to both university and college mission. That was followed by setting the students outcome based on an ABET accreditation body and American Society of Civil Engineers. These SOs are expected to be gained by the students from the different courses constituting their curriculum, by attaining the prescribed Course Learning Outcomes (CLOs). The assessments and evaluations of PEOs, CLOs and SOs were essential components in the assessment of the whole program which leads to decisions and actions that improve the program. In this regard, they mapped CLOs of each course to the most relevant SOs- taking into consideration that the least number of mapped SOs with CLOs is the better- and documented that in a mapping table. It worth to mention that weighted numbers from 1 to 5 were used in the mapping, where the weights represent the strength of each CLO in fulfilling the SO. Moreover, the course credit hours are also taken into consideration to determine how strong the course contributes in attaining the relevant SOs. That was computed using the following formula:

$$\text{Credit hour assigned for each SO} = \left[ \frac{\text{total weights of each SO}}{\text{Sum of weights}} \right] \times \text{course credit hour} \quad (1)$$

Due to the ever-increasing importance of outcome- based models, proofed by demanding such model by many accreditation bodies such as: ABET, International Engineering Alliance (IEA) and National Commission of Academic Accreditation and Assessment (NCAAA) [6]- many research works were devoted to shed the light on success stories of adapting this model in assessing different academic programs and getting international accreditations. Some other works were focused on the implementation of a particular stage in the model. Following are some examples.

A complete framework has been proposed in 2019 [7], [8] which is built on set of sustainable practices to advance quality assurance on outcome-based education systems and ensure delivering effective academic program by higher education institutions. It is worth to mention that this framework guided their four programs to successful ABET accreditation. The framework consists of four main types of activities which are: strategic planning, educational practices and strategies, assessment and evaluation and continuous improvement. These activities lead to strategically aligning the program outcomes with the market needs, delivering relevant skills through educational practices, evaluating the attainment level of graduating students and defining remedial actions to improve attainment, in case any deviations in performance were observed in the assessment results.

This framework, as recommended by many international bodies such as ABET with respect to student outcome centered approaches, requires setting measurable learning objectives of the academic programs, measuring those objectives with appropriate assessments to assess and evaluate the effectiveness of the academic programs, as well as having a continuous improvement process to identify the weaknesses in the programs and rectify it quickly.

With respect to the assessment and evaluation phase, they try to make the process comprehensive and at the same time lightweight to avoid extra load on faculty. They used a

combination of direct and indirect assessment methods. The direct assessment was through faculty members and the taught courses as well as an exit exam. With respect to courses assessments, SOs were assessed through formative assessments in the entry level courses and summative assessments in the exit level courses. The formative assessments gave warning signs indicating that a problem is there which needs to be resolved to improve the attainment level. Furthermore, faculty are informed at the beginning of the semester whether their course assessments will be used for program assessments to aid in planning their courses assessments. Courses to be covered in program assessments should have standardized rubrics to assess the attainment of the performance indicators. Such rubrics helped faculty in reducing the variations in assessment process. A program quality assurance team collaborates with faculty to assure that assessment questions intended to assess a particular SO are in line with the rubric. With respect to the exit exam, students were required to do an exam at the time of their graduation, where each question was designed for assessing a specific performance indicator. However, the students did not solve the exam seriously as it did not have any academic weightage. Thus, it was decided not to use the exit exam as an assessment tool again. Indirect assessments allowed them to hear the voice of other stakeholders through surveys and meetings with external advisory board.

Jalil in 2019 [9] shared the experience of Energy and Renewable Energy Program in Electromechanical Engineering Department in University of Technology-Iraq in measuring SO attainment while preparing for ABET accreditation. He detailed the process of measuring SO attainment for the four years' work in the program. Although, author frankly stated that attainment of student outcomes is a challenging task for all Iraqi Universities, ABET approved the adopted process as an appropriate way of measuring SO attainment. The Electromechanical Engineering Department started adopting the process of SO assessment in 2014, in which it focused on final exam grades in the assessment and evaluation. After two years, it changed the policy, after consulting ABET experts, and started considering the grades of different course assessments. To achieve that the department put the assessment plans; covering assessments methods, their mapping to CLOs, and the mapping between CLOs and SOs; in the course syllabus at the beginning of the year. Here it is worth mentioning that the department adopted many to many mapping between CLOs and SOs. Faculty members shall prepare the prescribed course assessments with detailed rubrics and performance indicators. Moreover, faculty members shall decide the target to be achieved by the student in order to consider the SO attained for each course, taking into consideration that target shall be above 60% since 50% is the pass degree set by the ministry. After grading course assessments, the average grade of the student is compared with the target and only if it is above the target, the corresponding SO is considered met. In case of an SO not attained, an action plan should be developed to improve the attainment. At the end of the year, the department prepares assessment reports - detailing assessment results, visual graphs, and the summary- and uses these reports in the annual review. The analysis over the year depicted that SO attainment was improved when they relied on different assessments with detailed rubrics rather than

relying on final exam grades only. Moreover, adopting problem-based learning by many faculty members in different courses helped in improving the attainment results. However, it was mentioned that attainment level still needs to be improved in order to qualify the program for ABET accreditation.

In [6], one of the stages of an outcome-based model was spotlighted, which is data preparation. The author presented a coherent vision of a systematic approach of data preparation that aids in collecting the required data (i.e. assessments' data that are mainly used for SOs evaluation as well as program's data that talk about the state of affairs of multiple stakeholders such as students, faculty members, facilities and institute) needed for the accreditation visit by different accreditation agencies. His approach started by identifying the different tasks; which were 23 tasks, to be accomplished in order to satisfy ABET requirements. That was followed by mapping each task to the required resources/ committees, from 7 well defined resources, and defining an interface between the different resources, which define explicit communication between their inputs and outputs. The effectiveness of this approach was illustrated by preparing data- to be used for accreditation- that revealed the real image of the program. Furthermore, it increases awareness, among faculty members, of the importance of such stage in conducting a successful assessment of a program as well as their roles and responsibilities. This led to reducing the load on the program coordinator and smoothed the process of getting the required data from all the faculty members.

Another effort was made in [3] that focused on the design and implementation of performance metrics for successful evaluation of assessments data. This work was made due to the importance of such performance metrics in quantifying the achievement of both program objectives (at program level) and student outcomes (at both program and course levels). The performance metrics were attainment, student achievement and x-th percentile. These metrics with the formulas behind them were tested and their results were analyzed on a sample course. The analysis revealed that getting high values of one metric did not lead to high values of the other metrics and their values depended on the scored marks distribution. They generalize their findings by stating that students' achievement is often less than SOs attainment for mean value which is less than the average marks that is in turn less than the passing threshold for uniformly distributed marks.

## B. Summary and Research GAP

The following Table I summarizes the literature review and provides a critique to each of the research works mentioned above:

The above summary indicates that there is a need for a framework that is not subject to recall bias from its main constituencies. Moreover, it is desired to be based on SOs assessments and evaluation to facilitate the accreditation by many international accreditation bodies. Furthermore, it shall have a systematic approach of aggregating data from different sources. In addition, should provide a comprehensive assessment to the entire program's SOs using a wide range of assessment tools, such as: courses' assessments, surveys, major projects, just to mention a few. Added to that, the framework

shall have a precise mean of computing the attainment level of each SO, to be compared with an absolute metric goal that is decided by the offering college / department based on the

collective experience of multiple faculty members and the market need.

TABLE I. LITERATURE REVIEW SUMMARY

Author , Date	Assessment and Evaluation Model	Main Focus	Results	Our Critique
Plaza et al., n.d., 2007	Curriculum mapping technique	-The intention is to verify that intended, delivered, and received curriculums are the same based on outcomes statements. -Intended curriculum is examined through the courses' syllabi and materials prepared by faculty members, the taught curriculum is examined through interviews with faculty members teaching/ coordinating the courses, and the received curriculum is explored through focus group interviews with students.	Concordance between the intended/ delivered and received curriculums	-This assessment model is not widely adopted since institutes are looking for international accreditations and the majority of the accreditation bodies require student outcome based models to be implemented. -This model is subject to recall bias from program constituencies (i.e. faculty members and students)
Iqbal Khan et al., 2016	Student outcome based model	-It is a comprehensive work detailing the success story of getting the Civil Engineering program accredited by ABET -It has been applied on the whole curriculum, considering the least mapping between SO and CLO. -Weighted mapping between SO and CLO was applied. -Courses were assigned rates regarding their impact in achieving the SO and their credit hours was considered in finding the attainment.	The proposed procedure is applicable for developing a new engineering program or adjusting an existing one to serve both the college and university missions and goals.	This paper provides a precise mathematical formula for computing the attainment level of each SO
Almuhaideb & Saeed, 2020	Student outcome based model	A full framework qualifying 4 computing and engineering programs to be accredited. -Lightweight (to avoid extra load on faculty) and comprehensive process (Applied on all courses as well as exit exam). This was achieved by treating assessments in entry level courses as formative and assessments in the exit level courses as summative one.	The lightweight assessment process was effective in assessing all the program SOs. However, it was noticed that Exit exam was not a reliable assessment tool since there are no academic weightage on it and students do not solve it seriously. Thus, future evaluation process will be limited to courses' assessments.	This work provides a comprehensive but lightweight assessment to all SOs of the program
Jalil, 2019	Student outcome based model	A process followed by Energy and Renewable Energy Program in adopting SO based model to be eligible for ABET accreditation. -In 2014, started adopting SO for evaluation on final exam grades only but in 2016 they started considering the grades of different course assessments. -It adopted many to many mapping between CLOs and SOs. -The metric goal is different from one course to another determined by the instructor (but should be above 60%).	SO attainment was improved when they relied on different assessments with detailed rubrics rather than relying on final exam grades only. Moreover, adopting problem-based learning by many faculty members in different courses helped in improving the attainment results. However, it was mentioned that attainment level still needs to be improved in order to qualify the program for ABET accreditation	-Metric goal should not be determined by an individual (i.e. faculty member) but would be better to be determined by the offering department or college taking into consideration different faculty opinions and experience to reduce the possibilities of human errors.
Rashid, n.d., 2021	Student outcome based model	Focus on one stage of student outcome based model which is data preparation by identifying required tasks and mapping responsible committees or needed resources to them.	Ease in collecting data that revealed the real image of the program from faculty members and reduce the load on the program coordinator	A great effort in coming up with a systematic approach of aggregating the data from different sources smoothly. This general approach could be used by any institute while collecting and preparing the data to be used for accreditation purpose
Ahmed & Bhatti, 2016	Student outcome based model	Design and implementation of performance metrics for evaluating the assessments data. These were attainment, student achievement and x-th percentile. -The proposed passing threshold is the min (70% , avg). -The study was applied on sample courses and not all the courses in the curriculum.	getting high values of one metric did not lead to high values of the other metrics and their values depended on the scored marks distribution. They generalize their findings by stating that students' achievement is often less than SOs attainment for mean value which is less than the average marks that is in turn less than the passing threshold for uniformly distributed marks	Passing threshold should not be relative to the students score. It should be an absolute value determined to ensure that attaining the student outcome will lead to satisfying the market need.

III. PROPOSED FRAMEWORK

A. Development of PEOs and their Mapping to SOs and CILOs

The outcome-based assessment framework proposed started by defining the Program Educational Objectives (PEOs) for the BSIT program, taking into consideration that educational program must be regularly reviewed, assessed, and evaluated [7], [10] for many reasons, including:

- 1) Meeting educational standards and measures of quality: This award the program a recognition of its quality, integrity and performance which lead to receiving confidence from educational community and the public [5].
- 2) Ensuring that PEOs are based on the program stakeholders needs as well as assessing the degree of attaining these PEOs which is an important sign of the level of success achieved [5].
- 3) Improving student learning experience.
- 4) Ensuring the sustainability of the program [11].

The final list of PEOs was discussed and approved by the main program constituencies (i.e. faculty, alumni, external advisory board, and employers); which is:

- 1) Exhibit the relevant skills and knowledge for pursuing the IT career in industry including corporate as well as government sector.
- 2) Pursue life-long learning leading to entrepreneurship, research and development.
- 3) Contribute to the society through their ethical and professional norms by demonstrating them in IT professions.

These PEOs are supposed to be fulfilled through the different courses in the program curriculum. Table II presents the mapping of ABET six SOs to BSIT three PEOs.

Each course has a set of CILOs that should be attained by the students upon completion of the course. Each student outcome is translated to the relevant performance indicators to assess the student outcomes within a program. To assess the student outcomes for the program, each performance indicator is mapped to the relevant CILOs of that particular course. Table III shows the complete view of the BSIT program student outcomes in relation to performance indicators as well as their one-to-many mapping to CILOs, assuming equal weights across all CILOs mapped to a particular SO.

TABLE II. BSIT SOS- PEOs MAPPING

PEOs \ SOs	1	2	3
1	√	√	
2	√	√	
3			√
4	√		√
5			√
6	√	√	

TABLE III. STUDENT OUTCOMES AND PERFORMANCE INDICATORS FOR IT PROGRAM

Student Outcomes	Description	Performance Indicators	CILOs
1	Analyse a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions.	P1.1- Understand computing concepts to solve complex problems	A1
		P1.2- Understand the latest trends and technologies.	A2
		P1.3- Analyse complex computing problems	C1
2	Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline.	P2.1- Model a computing-based solution	B2
		P2.2- Design a computing-based solution	B2
		P2.3- Create an innovative computing-based solution	C3
		P2.4- Integrate various components to the computing-based solution	C2
		P2.5- Implementing a computing-based solution using appropriate tools	B3
		P2.6- Evaluate a computing-based solution	C1
3	Communicate effectively in a variety of professional contexts.	P3.1- Use of appropriate methods/tools for communication	D1
		P3.2- Explain the ideas effectively in written	D1
		P3.3- Organize information properly in written	D3
		P3.4- Present information effectively and well organized	D3
		P3.5- Present the ideas appropriately in front of variety of audience	D1
4	Recognize professional responsibilities and make informed judgments in computing practice based on legal and ethical principles.	P4.1- Understand professional practices / standards	A3
		P4.2- Understand legal and ethical principles	D4
		P4.3- Make ethical and professional Judgements in computing practices	D4
5	Function effectively as a member or leader of a team engaged in activities appropriate to the program's discipline.	P5.1- Perform individual task to meet team goals	D2
		P5.2- Work effectively with other team members	D2
		P5.3- Act as an effective leader to other team members	D2
		P5.4- Plan the tasks and resources to develop computing-based solutions	D3
6	Apply computer science theory and software development fundamentals to produce computing-based solutions.	P6.1- Understand the problem domain	B1
		P6.2- Identify user needs within the given problem	B1
		P6.3- Apply software development fundamentals in developing computing-based systems	B1

## B. Assessment and Evaluation Process of PEOs, SOs and CILOs

A number of data collection methods as well as assessment tools have been defined to facilitate measuring the achievement of both CILOs and SOs and consequently PEOs. After that comes the stage of utilizing the results and feedback in continuously improving the program with its learning and assessment processes and achieving high quality standards.

The proposed approach has short term (every academic semester) and long term (every three years) processes. It is carried through three main stages, which are:

1) *Data collection and preparation:* Data is gathered from various potential- internal and external- sources that cover all the program stakeholders. The primary sources are:

a) *Faculty developed assessment:* A faculty can use variety of summative and formative assessment to evaluate the student during the semester. The grades of each question in the summative assessments need to be recorded for all the enrolled students in a course. This is done by filling the CILOs Evaluation excel sheet by the faculty member. This CILOs Evaluation excel sheet needs to be submitted as part of the course portfolio to the college by the end of the semester.

b) *Internship feedback:* each student is required to go through industrial training program. During internship, the assessment of the students is done by their industrial – site- and academic supervisors. The site supervisor evaluates the student twice; after the completion of the first month of the internship period and after the completion of the internship period. The evaluation is done by filling the mid-evaluation written questionnaire form and final evaluation written questionnaire form. The mid-evaluation form consists of 10 questions; with 5-points scale; and a section to write comments. The final evaluation form has the same set of ten questions plus a space for the evaluator to comment on the strength / weaknesses of the intern as well as provide recommendations on preparing students for the workplace. In similar way, the academic supervisor evaluates the student after the completion of the internship period by filling a written questionnaire form with 5-points scale. Furthermore, a summary report of all the student undertaking the industrial training will be prepared by the college internship coordinator and submitted to the college by the end of the semester.

c) *Major project examination and exhibition:* Students in their major project course are examined orally after submitting their written reports. The reports and oral exam aids in performance evaluation by the examination committee, which consists of two faculty members and an external examiner from the industry. Ready assessment forms are prepared to help in assessing and evaluating the students. Moreover, an exhibition is arranged to give the students an opportunity to present their developed projects to various businesses. Many people, including experts from IT field, attend the exhibition and spend ample amount of time in examining and evaluating the projects as well as the students' skills. Their evaluations and comments will be collected through a survey prepared by the college.

d) *Written Surveys:* The College collects feedback from variety of stakeholders and makes use of that in assessing and making informed decision. A third party is responsible for conducting the written surveys, collection and collation of multiple stakeholders' feedback that are passed for the colleges for further action. The main written surveys are:

- i) *Alumni survey:* This felicitates graduate's satisfaction with the program and their learning experience.
- ii) *Senior exit survey:* This reveals students' satisfaction level with respect advising, teaching and learning, administrative support and facilities. Moreover, it prompts students to overall rate the program.
- iii) *Employers' satisfaction survey:* This enables the college in receive employers of the graduates' feedback regarding graduates' job performance, thinking skills and character.

2) *Data collection and preparation:* The success of the program is assessed by the level to which it satisfies the university and college missions and goals. Since the PEOs were derived from the college mission which is in line with the university mission, assessing the PEOs leads to assessing the whole program. PEOs are assessed through the SOs and CILOs mapped to them.

This assessment process is carried at both course and program levels via various assessment tools. Some of these tools are directly assessing SOs and CILOS while others are indirectly assessing them. In case of indirect assessments, a mapping should be done to facilitate assessing CILOs and SOs. The level of CILOs/SOs attainment is compared with a metric goal set by the college. If the attainment level is below the metric goal, then it will be considered as not attained.

These assessment tools are:

a) *Faculty developed assessment:* It is a direct assessment tool carried by faculty teaching courses comprised by the curriculum. A Faculty member assesses students' attainment of CILOs, and in turn SOs, through different course assessments. The faculty will be recording the students' grades in CILOs/ SOs evaluation excel sheet for each course. This sheet details the summative assessment methods used in assessing the student outcomes. Because this serves as quantitative analysis, questions are used as the basic units of computation of assessment. That's why, CILOs/ SOs evaluation sheet records all the students' grade for each question in each assessment and provides a clear mapping between each question in the course assessments and CILOs/SOs. Here, it is worth to mention that each question could be mapped to maximum two CILOs/ SOs. In addition, CILOs/ SOs evaluation sheet provides a summary of the results, indicates the level to which each of the CILO/ student outcomes is being attained, and determines whether the CILOs/ student outcomes have been met or not by comparing their average percentages with the predefined metric goal. This leads to absolute student achievement evaluation since the metric goals is fixed to a target value. It is noteworthy that in case of having students from other academic programs in the courses,

separation must be done during the assessment phase to ensure that only IT program's students are considered in the assessment and evaluation.

*b) Internship written questionnaires and forms:* This is a direct assessment tool since the CILOs assessed by each evaluation form / submitted report is clearly written in the internship course syllabus. It worth to mention that 50% of the student final grade is taken from the total score given by the site supervisor in the two evaluations. However, only 10% of the score given by the academic supervisor will be counted. The remaining 40% of the student grade is based on the evaluation of the three reports submitted by the student during his internship period and after the completion of the internship. Furthermore, CILOs/SOs evaluation excel sheet will be prepared for this course to be taken into consideration in the program evaluation process.

*c) Written surveys:* These are indirect assessment methods. In alumni survey, "alumni experience with the program" section was mapped to SOs and thus used in the

assessment. In senior exit survey, "program overall rating" section was mapped to SOs and used in the assessment. In employer satisfaction survey, SOs were clearly listed to seek employers feedback on them. In addition, "employee learning outcomes and skills" section was mapped to SOs to aid in getting comprehensive assessment.

The aforementioned assessment tools are carried out with different frequencies. Every academic semester, all the direct assessments will be carried out in addition to the senior exit survey. Other surveys are conducted every year.

At the end of each academic year, the result of all the assessment methods will be compiled together, as shown in Fig. 1. That is done by giving weight to each of the assessment tool. In this regard, each of the survey was given 10% and by that we assure balance consideration of the stakeholders (students, alumni and employers) and 70% was given to faculty developed assessments which represent the main direct assessment method. This will give us the percentage of attaining SOs in the year.

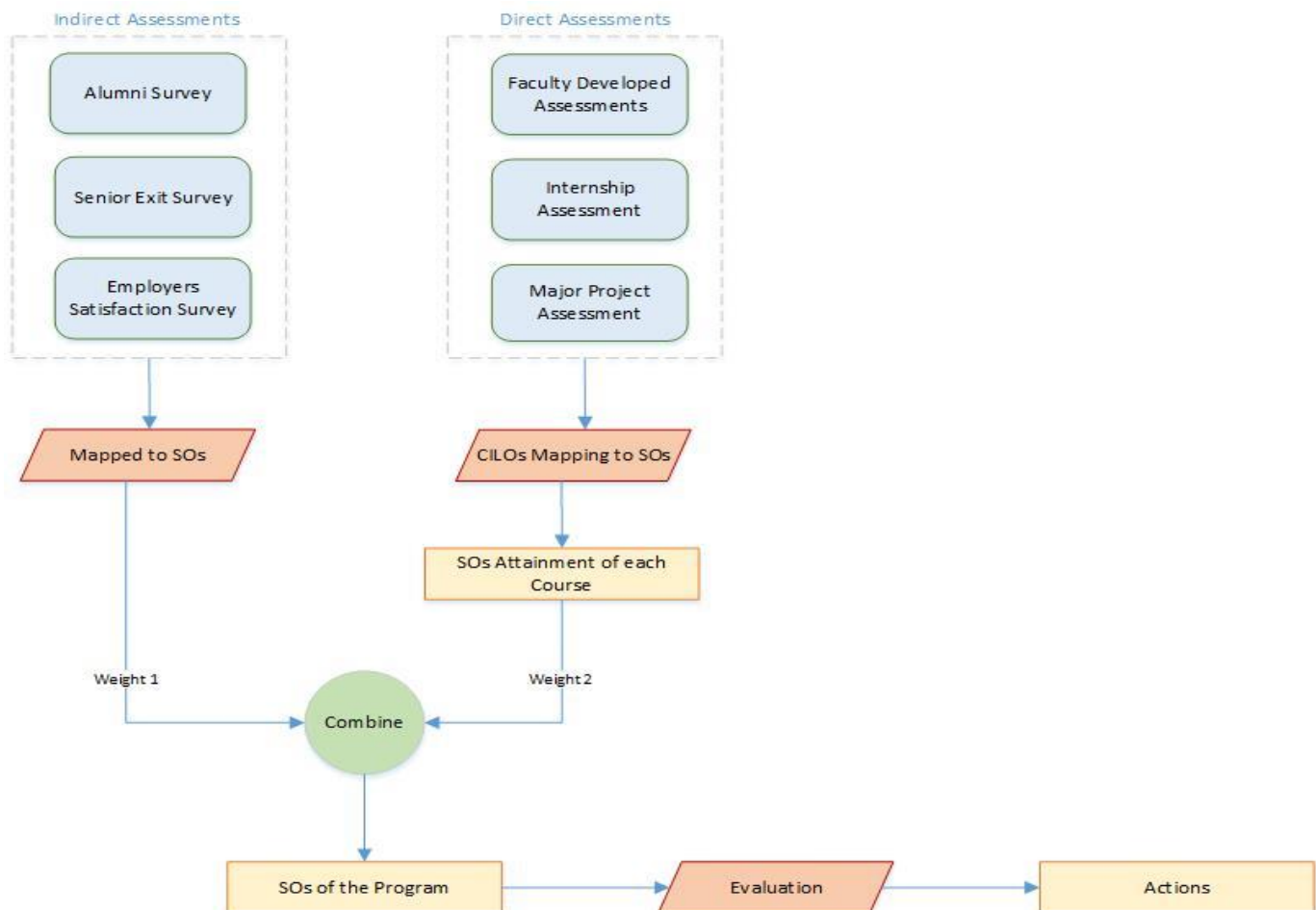


Fig. 1. Outcome based Framework.

3) *Evaluation*: At the end of each academic semester, IT College holds meetings to discuss thoroughly the assessment results, including: SOs achievements and surveys summary reports. With regards to SOs achievements, it gets evaluated on course level as well as on program level. In case of course level evaluation, the attainment of each SO is based on the attainment of CILOs that represent PIs mapped to it. An SO/CILO is considered attained, if the average of the students' grades of questions mapped to SO/CILO is not less than the predefined, absolute, metric goal which is 70%. If any of the SOs is not attained, the course coordinator/ instructor will be requested to provide a reason behind this pitfall and formulate an action plan to be carried on the next semester to improve the SO attainment level and overcome this downfall. On the next semester in which the course is offered, the IT College quality coordinator will follow-up with the faculty member to check the attainment level of those SOs that were not attained earlier. If any SO is still not attained, the reason should be provided as well as another action plan to be placed to be implemented on the following semester. Table IV presents some examples of SOs not attained on the course level and the process of monitoring them by the college. This practice had high impact on improving the teaching methods since faculty members were encouraged to evaluate the effectiveness of their teaching and adopt new techniques that were not used previously in an aim to improve SOs attainment. Moreover, some faculty tried to boost the attainment by revising and adjusting the lesson plan and devoting more time for topics/ concepts that were not comprehended by students. In like manner, assessment methods had been improved because faculty members were motivated to use verity of assessments in assessing each of the SOs. This ensures getting reliable assessment results, because, even if one of the course assessments was not effective in assessing a particular SO, that will be overcome by other course assessments.

The same process is applicable for evaluating SOs on program level; by considering all the courses with CILOs corresponding to its PIs.

Achieving good results demonstrates that the PEOs are based on the needs of the program's various constituencies. In addition, it demonstrates the high competence of the program graduates in their professional career as well as the high level of their employer satisfaction. All the assessments and evaluations results and findings are well documented in the college meeting minutes as well as the annual report produced and maintained by IT College. The report clearly states the pitfalls/ deficiencies, justification of each, areas to be improved, action to be taken in order to improve the program outcomes, meet high quality international standards and ensure its sustainability. This aids in planning how to overcome the existing deficiencies in the near future. Such documentation and reporting reflect the ongoing evaluation in place leading to continuously improve the program effectiveness.

#### IV. RESULTS AND DISCUSSION

The longitudinal methodology opted in this research enabled in assessing SOs achievement and evaluating BSIT program over the last five years. Data was first collected at the outset of the study and then gathered, repeatedly, during the course of the study (i.e. five years). That facilitates observing changes in SOs achievements over time.

Fig. 2 provides empirical insights of the performance of the College of IT after adopting the proposed framework. It summarizes the results of the assessments approach over 5 academic semesters. It illustrates the extent to which each of the SOs is being attained using the weighted combination of direct and indirect assessment methods.

The student outcomes achievements were maintained high over the last five semesters. This comes as a result of the careful implementation of the proposed framework that enabled the College of IT in having a systematic assessment and evaluation process, leading to achieving outstanding results in quality assurance and being ABET accredited. Moreover, there are number of notable routines and practices that did support the presented framework in achieving these satisfying results. Among these practices are advising and counseling that provide customized and personalized advice to individual students guiding them in defining and achieving their academic, career and life goals. That enabled the college in maintaining good student ratio. Furthermore, different workshops are regularly conducted by the college to qualify faculty members with new teaching and assessment methods. Added to that, reviewing the courses' syllabi, routinely, at the beginning of every academic semester and ensuring having up-to-date textbooks brings new advancements in IT into the IT curriculum and confirms that the program is up to date.

The results of the implementation proved the effectiveness of the proposed framework together with other educational practices in improving the student performance and the program.

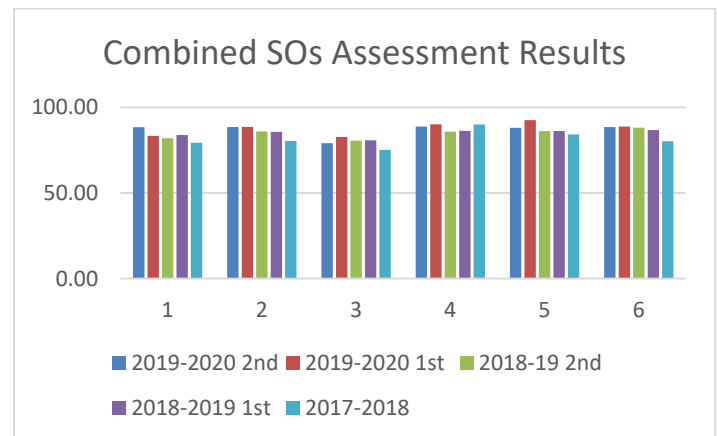


Fig. 2. Student Outcomes Attainment

TABLE IV. EXAMPLES OF STUDENT OUTCOMES ACHIEVEMENT ANALYSIS

Course	SO	Reported Semester	Current Semester	Instructor	Met/	If Not, Specify the reason	Improvement Plan	Actions Taken	Previous Achievement	Current Semester Achievement
					Not Met					
ITCS 404	1	1st 18-19	1st 19-20	Instructor1	Not Met	More focus was on the solving problems + Poor in critical analysis	The theoretical background is to be discussed more in the class + More exercises and discussion	class discussion + More exercises are given in the class	51.48	77.98
	2									
ITCS 209	1	1st 19/20	2nd 19-20	Instructor2	Not Met	more class discussion was needed + weak math background of the students	More time will be devoted to class discussion + More exercises will be given through Kahoot	More class discussion and exercises were given	61.9	84.45
	6									

Getting such outstanding results does not mean that the IT program does not need continuous improvements. In this regard, the College of IT believes that curriculum development is a dynamic process that ensures having an up-to-date program in line with the rapid advancements in Information Technologies and in compliance with ABET requirements in case ABET changes any of their criteria or requirements.

#### V. CHALLENGES

The College of IT faced number of challenges during the course of implementing the proposed framework. Following is the list of challenges together with how did the college overcome each:

##### A. Balancing between NQF Requirements and ABET Requirements

Bahrain National Qualification Framework has a set of standards and requirements to be satisfied by an academic program in order to endorse its associate qualification and be placed on a specific NQF level (from level 1 for Access1 to level 10 for Doctoral Degree).

BSIT satisfies the standards set by NQF of level 8 [12]. However, in order for BSIT to be accredited by ABET, the program should also satisfy the requirements of ABET. To achieve that, the college came up with a mapping of PEOs, SOs and CILOs, as shown in Table I and Table II.

##### B. Having Students from Different Majors in BSIT Program Courses

Some of BSIT program courses are offered for students from other colleges (such as Engineering College and Business College). However, calculating the attainment of SOs for BSIT program requires considering IT students only. In this regards, the college decided to separate the result of IT students from other students and simply consider IT students' grades in the assessment and evaluation. In case of having few IT students

(i.e. less than four) registered in the course, the results will not be considered in the evaluation of the program.

##### C. Online Course Delivery during COVID-19 Period

Although classes were delivered online during the pandemic, but all the courses content and CILOs were covered and assessed. Even the practical components of the courses were delivered through virtual labs. Thus, the process of assessment and evaluation of SOs were not affected.

##### D. Market Needs and ABET Requirements

Market needs were analyzed and considered while designing BSIT program as well as formulating its goals and objectives. BSIT graduates should be qualified enough to fulfill market needs by having distinguished and desirable skills and knowledge. Fortunately, the proposed framework secured the alignment of program outcomes with market needs by ensuring delivering the desired skills, evaluating the attainment of graduates and taking a remedial action in case any of the SOs was not attained.

##### E. Effect of the Correctness of Mapping of SOs to CILOs on the Reliability of the Evaluation Results

It is well known about SO based model that the quality of its evaluation result depends on the quality of the mapping between CILOs and SOs. Thus, program coordinator (who creates the mapping between CLOs and SO) and instructors (who map assessments' questions to CILOs and SOs) shall be qualified enough to do correct mapping to ensure having reliable results at the end. Due to the highly qualified faculty members in the College of IT in Ahlia University, the obtained results were reliable and trustworthy.

##### F. Overloading Faculty Members during the Data Collection and Assessment Phases

Faculty were overwhelmed with their responsibilities in the data collection and assessment phases. However, it has been



planned to reduce that in the near future and have a lightweight process by selecting some of the courses and their assessments to be considered in the program assessment and evaluation, while other course assessments will be considered as formative one. This should reduce the load on faculty while maintaining good evaluation result that aids in informed decision making and continuous improvement of the program.

## VI. CONCLUSION

The student outcome-based framework presented in this paper has been adopted by College of IT in Ahlia University for the last five years. The vast amount of data collected by the college during this period, boosts the robustness of this framework in extracting continuous improvement requirements. Moreover, number of assessment tools were in use to measure the achievement of CIOs, SOs and in turn PEOs and their results were utilized in making informed decisions regarding continuously improving the program with its learning and assessment processes, achieving high quality standards as well as assuring the sustainability of high quality education.

Although some challenges were faced during the implementation of the proposed framework, this framework helped the college in improving the academic program. Other educational practices supported this framework such as: academic advising and career counselling. The success is reflected by improving the curriculum through adding new courses, removing courses, and modifying the content of some courses based on the market needs, international standards, constituencies feedback and the new trends in IT. Moreover, enhancing current teaching and assessment methods as well as adopting new one is another sign of improving the program that promotes positive assessment results. Furthermore, achieving distinguished and honorable results- such as improving students' performance, attaining high level of graduate attributes, achieving high quality standards, being accredited by ABET and satisfying various stakeholders- are clear indications of the success and improvements in the program.

Despite the fact that the College of IT adopted SOs defined by ABET as SOs for BSIT program, the proposed framework will still be applicable even if ABET re-define their set of SOs. Moreover, it is also applicable to other academic programs.

## REFERENCES

- [1] "Criteria for Accrediting Computing Programs, 2020 – 2021 | ABET."
- [2] B. El-Ariss, A. M. I. Sweedan, and K. M. El-Sawy, "Civil Engineering Course Outcome Assessment," 2022. [Online]. Available: <https://www.researchgate.net/publication/242493700>.
- [3] I. Ahmed and A. Bhatti, "Design and Implementation of Performance Metrics for Evaluation of Assessments Data," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 5, no. 3, pp. 235–245, 2016, [Online]. Available: <http://iaesjournal.com/online/index.php/IJERE>.
- [4] C. M. Plaza, J. Reiersen Draugalis, M. K. Slack, G. H. Skrepnek, and K. A. Sauer, "Curriculum Mapping in Program Assessment and Evaluation," 2007.
- [5] M. Iqbal Khan, S. M. Mourad, and W. M. Zahid, "Developing and qualifying Civil Engineering Programs for ABET accreditation," *Journal of King Saud University - Engineering Sciences*, vol. 28, no. 1, pp. 1–11, Jan. 2016, doi: 10.1016/j.jksues.2014.09.001.
- [6] M. Rashid, "A Systematic Approach of Data Preparation for ABET Accreditation Design Space Exploration for 3D architectures View project Proofs A Systematic Approach of Data Preparation for ABET Accreditation\*," 2021. [Online]. Available: <https://www.researchgate.net/publication/349521664>.
- [7] A. M. Almuhaideb and S. Saeed, "Fostering sustainable quality assurance practices in outcome-based education: Lessons learned from abet accreditation process of computing programs," *Sustainability (Switzerland)*, vol. 12, no. 20, pp. 1–21, Oct. 2020, doi: 10.3390/su12208380.
- [8] A. Shafi, S. Saeed, Y. A. Bamarouf, S. Z. Iqbal, N. Min-Allah, and M. A. Alqahtani, "Student Outcomes Assessment Methodology for ABET Accreditation: A Case Study of Computer Science and Computer Information Systems Programs," *IEEE Access*, vol. 7, pp. 13653–13667, 2019, doi: 10.1109/ACCESS.2019.2894066.
- [9] J. Jalil, "Measurement of Student Outcomes Attainment for ABET Accreditation in Electromechanical Engineering Department in UOT," *Engineering and Technology Journal*, vol. 37, no. 4C, pp. 449–453, Dec. 2019, doi: 10.30684/etj.37.4c.12.
- [10] H. Goss, "Student Learning Outcomes Assessment in Higher Education and in Academic Libraries: A Review of the Literature," *The Journal of Academic Librarianship*, vol. 48, no. 2, p. 102485, Mar. 2022, doi: 10.1016/J.ACALIB.2021.102485.
- [11] A. Hidayah STAI Rasyidiyah Khalidiyah Amuntai Kalimantan Selatan and S. STAI Rasyidiyah Khalidiyah Amuntai Kalimantan Selatan, "Internal Quality Assurance System Of Education In Financing Standards and Assessment Standards," 2022.
- [12] "Higher Education Review Reports - Higher Education Review Reports." <https://www.bqa.gov.bh/En/Reports/UniReports/Pages/NewReport.aspx> (accessed Jan. 23, 2023).

# A Hybrid Filtering Technique of Digital Images in Multimedia Data Warehouses

Nermin Abdelhakim Othman<sup>1</sup>, Ahmed Ayman Saad<sup>2</sup>, Ahmed Sharaf Eldin<sup>3</sup>

Dept. Information Systems-Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt<sup>1</sup>

Dept. Information System-Faculty of Informatics and Computer Science, British University in Egypt, Cairo, Egypt<sup>1</sup>

Dept. Information Systems-Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt<sup>2</sup>

Dept. Information Systems-Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt<sup>3</sup>

Dean-Faculty of Information Technology and Computer Science, Sinai University, Sinai, Egypt<sup>3</sup>

**Abstract**—The similarity search approach used for image Data Warehouse (DW) can provide better insights into discovering the most similar images compared to the input query. Due to the later innovation improvement, the mixed media complexity is discernibly expanded and modern inquires about regions are opened depending on comparable mixed media substance recovery. Content-Based Image Retrieval (CBIR) algorithms are utilized for the retrieval of images related to the inquiry image from gigantic databases or DW. The queries that are used for DW are complex, take a lot of time to process and many give less accurate results. For these reasons, this paper needs to have an effective technique to improve the similarity search query process that reflects a more positive result. In this paper, show how to extract features from a set of images (color, shape, and texture features) by using CBIR algorithm with Color Edge Detection (CED) method. Once these features are extracted, the proposed method will minimize the distance between these features vectors and the query image one using a Genetic Algorithm (GA). This paper illustrates the extraction of endless strong and imperative features from the database of the images, therefore, the capacity of these features in storing within the frame of features vectors. Accordingly, an imaginative closeness assessment with a metaheuristic algorithm (Genetic Algorithm (GA) with Simulating Annealing (SA)) has been attained between the query image features and those having a place in the database image. This paper introduces a new algorithm CEDF (Color Edge Detection with Gaussian Blur Filter) that applies the Gaussian Blur Filter after using CED method for feature detection of the image. Experimental results show that CEDF method gives better result than the other already-known methods.

**Keywords**—Data Warehouse (DW); Content-Based Image Retrieval (CBIR); Color Edge Detection (CED); Genetic Algorithm (GA); Simulating Annealing (SA); Memetic Algorithm (MA)

## I. INTRODUCTION

A perfect image processing system relies on three vital factors; i.e. acquiring the image, transmitting the acquired image, and stretching security to the image through its transmission. This leads to defective examination of medical images while acquiring an image from the database. Therefore, this paper needs an effective method to remedy this corruption. We present a new method called CEDF. This method added a new value to the image processing process Fig. 1.

Hence, the perfect image processing system consists of three important phases: Image de-noising, image compression and image security.

### A. Image Filtering

Image filtering is considered a very important phase within the image processing system. The image captured by the camera is debased by noise. That is loud image may degenerate more through the transmission process. Such a noisy image may regularly obstruct the operation of the communication system. Subsequently, it is exceptionally basic to have an effective de-noising method to evacuate noise from the image. As per noise pixel dissemination values, motivation noise is composed of two types. These types are irregular and fixed esteem motivation commotion. This commotion is called salt and pepper noise. As the noisy pixel can either have the most elevated or lowest value on grayscale [1].

### B. Image Compression

As with image de-nosing, image compression is additionally an essential stage of the image processing system. The transmission and storage of crude images request huge memory space. A great compression technique with quicker processing and memory-productive compression ability fulfills the requirements of the cutting-edge image processing system. In common, compression leads to the compression of information within the advanced image. The best objective of the compression procedure is to eliminate the repetition of image information. Therefore, the computerized image will be put away more viably. Lossless and Lossy compression techniques are the two essential categories of compression procedures. In the lossless compression strategy, an image sometimes before and after compression is identical to each other. Moreover, each bit of image is put away in the decompression process. While, the lossy compression procedure, there is a contrast between the original and reproduced image. But the reconstructed image is sensibly closer to the initial image [1].

### C. Image Security

The rise of the web causes extraordinary danger to data. These data are stolen through its transmission over the communication channel. Hence, the security of image

information is additionally one of the very crucial perspectives of the image processing system. Cryptography is one of the strategies utilized to supply security to data. In addition, parcel of other methods is developed to supply security to information. Sometimes parcel is not conceivable to keep the message content secured. In such a case, it is basic to keep the presence of the message as a secret. This strategy of giving security to a message is called steganography. Steganography is diverse from cryptography with reality. Cryptography bargains with keeping the message content emitted while steganography bargains with keeping the existence of the message secreta [1].

The bit-flipping mutation is a popular operator, in which a single bit in the string is flipped to form a new offspring string. A variety of other operators has also been developed, but is used less frequently (e.g., inversion, in which a subsequence in the bit string is reversed). A primary distinction that may be made between the various operators is whether they introduce any new information into the population. Crossover, for example, does not while mutation does [2].

All operators are also constrained to manipulate the string in a manner consistent with the structural interpretation of genes. For example, two genes at the same location on two strings may be swapped between parents, but not combined based on their values [3].

In this paper, the extraction of features (color, shape and texture features) will be shown from a set of images. Once these features are extracted, the method will minimize the distance between these features vectors and the query image one using GA. The paper carried out a performance test over an image dataset. Experimental result shows that CEDF results in a recall and precision raised average by 1.4 and 8.05 respectively, which assess and demonstrate the accuracy of the proposed algorithm.

This paper is organized as follows; Section II presents background related to the proposed method, Section III sheds some light on related work, Section IV speaks about the proposed algorithm, section, Section V deals with rating the performance of the proposed method, Section VI measure performance and evaluation, Section VII discusses the experiments and results, and finally, Section VIII is concluding the conclusion.

## II. BACKGROUND

This paper detail the most issues related to similarity search over images in Segment 2.1. Moreover, the proposed work is based on two well-known concepts used to use in all of the research. These concepts are talking about image features and are accessible within the literature. The Omni-technique and the star-join Bitmap index are portrayed in Segments 2.2 and 2.3, separately.

### A. Similarity Search

To be computationally analyzed, images ought to be pre-processed utilizing include extractors. These extractors are capable of generating features vectors that depict the

natural characteristics. This preparation is elaborated as takes after. An image is spoken to as a two-dimensional  $m \times n$  matrix of pixels. Where  $m$  and  $n$  are the image measurements and the pixel has numbers of values that depend on the image sort. For occurrence, the pixel values can be 0 or 1 in parallel images shift between 0 and 255 in grayscale images spoken to by 8 bits and have their values within the run of 0–255 each in RGB color images. A feature descriptor is characterized by: (i) a feature extractor calculation, which tracks down the images. That feature processes their pixels values and produces number of distortions of them, and stores these values in included vectors. (ii) A distance function, which produces a similarity measure. That is utilized to decide at that point the divergence between two images based on their features vectors.

The natural characteristics of images are as a rule depicted by properties with respect to color, texture, and shape. For occurrence, feature extractors may be actualized to calculate Color Histograms [3], the Haralick descriptors [4], and the Zernike moments [3]. Color histograms speak to the conveyance of colors (or levels of gray) in an image by calculating the recurrence. In which the color intensity of each pixel happens. With respect to texture feature extractors, the Haralick descriptors are employment factual approaches. That is to determine co-occurrence networks of images such that each matrix speaks to the connection between the pixel's position and its values.

### B. The Omni Technique

The Omni technique [3] is based on the choice of agent images (i.e. foci) from the DB, which is strategically positioned within the metric space. The number of representative images is given by the inborn dimensionality of the DB whereas the situating is given by the Hull of Foci algorithm. The main thought behind this calculation is to select images near the DB's borders. That can be utilized to progress probability in similarity searches.

### C. The Star-join Bitmap Index

A fundamental Bitmap list [3] built on an attribute comprises of several bit-vectors. One for each esteem an of  $A$ , where the  $i$ -th bit is 1 in case the  $i$ -th push is rise to  $a$ . There is something else; the  $i$ -th bit is 0. In DW, a Bitmap list can be developed to list properties of the dimension tables demonstrating the set of tuples in a fact table. That table holds the corresponding quality values. A Bitmap list with such a plan is called the star-join Bitmap (SJB) index [3].

In this paper, the extraction of features from a set of images (color, shape, and texture features) will be cleared. That is happening by utilizing CBIR algorithms with Color Edge Detection (CED) method. Once these features are extracted, the method will minimize the removal between these features vectors and the query image one employing a genetic algorithm. This paper outlines the extraction of unending solid and basic features from the database of images. Moreover, the capacity of these features is inside the store of features vectors. In like manner, a creative closeness appraisal with a metaheuristic algorithm (GA

with SA) has been achieved between the QI features and those having a put to the database image.

### III. RELATED WORKS

The related work will be divided into three directions of research that is synchronizes with the three phases of de-noising the image. The next sections will conclude the illustration of the following points:

A. Content-Based Image Retrieval, B. Image Similarity-Based Measure and C. Image de-noising.

#### A. Content-Based Image Retrieval

All of the previous research is important and still under research. However, no one mentioned the image as a perfect part of the source of data. Including images is vital as it improves the performance of the DW and broadband the capabilities of DW. That supports more accurate decisions for decision-makers.

A “Trigger” algorithm is used to recognize the lateness in tracking the unusual results in a large national clinical data warehouse of electronic health record (EHR) data [5]. It uses a method that upon the available data in EHR data repository from all departments of veterans’ affairs healthcare facilities. This method analyzes data from seven facilities. The limitations of this research are developing and refining similar algorithms more widely. In addition, that can potentially reduce delays in diagnostic evaluation and improve the quality and safety of patient care.

In addition, there is a research depended on novel identification-based image correction method using a bi-illuminant dichromatic reflection model. That occurs by a previous awareness demonstrated by small patches with close reflection properties, called Spectrum-Shape Elements (SSE) [6]. They introduced a Hausdorff-like metric for the set of spectrum shape elements. The basic properties of these SSEs and their localization methods were proposed. The shortages of this paper are: 1) constructing SSE pairs representing prior knowledge about distortions, 2) distortion-free illumination, 3) adapting the correction technique to other image distortion sources. This technique is showing the sensor noise and pixel sampling.

The author in [7] deals with textual and non-textual features to retrieve similar images by using a novel CBIR algorithm. After that, the algorithm will classify the query image into textual or non-textual categories. In the case of the textual query images, the text is revealed and realized. For the case of visual query, outstanding qualities are detected. Both qualities merged to shape the final feature vector. In addition, through these features, the top-ranked images are retrieved.

Besides this approach is effective for textual and non-textual images, its contribution is as follows: 1) It can deal with three methods of retrievals. 2) It can deal with low-power hardware independent of training efforts. In [8] the paper is about fusing the visual and semantic similarities between the query image and the database image. Then, they use the shortest path algorithm to a weighted graph.

That is for every database image node that has its pairwise similarity measure. This method was conducted on the retrieval of the common CT Imaging of Lung Diseases (CISLs). Furthermore, the results are not only better but also were more efficient. The limitations of this paper are: 1) This method can be only used in a small database but for a large database. In addition, it will consume more runtime and that requires clustering for the images. 2) This method can be only used for single-label data.

The author in [9], is proposed a new idea for the efficient retrieval of top-ranked images. That is done in the available DBs depending on local image features and spatial information in BoW architecture. To detect the performance evaluation of the proposed method, it was performed on seven state-of-the-art detectors and descriptors. After conducting this method over ten different DBs, it showed the accuracy of the proposed method. Additionally, is cleared by high precision and recall percentage record. It also proved that the proposed method exceeds the state-of-the-art detectors. As well as exceeds the descriptors over color, texture, scattered objects and complex DBs. The drawback of that method is never considering the technique scale-invariant and noise-robust.

A new content-based histopathological image retrieval approach has been proposed. This approach proposed a multi-scale and multichannel decoder based on LTP. That is to get special characteristics for every single image histogram representation. It also conducted the VLAD coding on the resultant features. Besides, power-law normalization has been performed to shape the final feature vector. By comparing the state-of-the-art techniques with the proposed approach, it shows the rising retrieval accuracy of the proposed approach. Efficiency and computational complexity are the limitations of the proposed approach [10].

This framework chooses the top-ranked images features. These features are most similar to the query image features for image retrieval. The main goal of any CBIR algorithm is to calculate the two basic descriptors (mathematical features and physical features). The main difficulties that are faced these methods are: 1) minimizing the semantic gap between the results and user expectations, 2) caring about low-level image descriptors [11].

The author in [12], is proposed a new method of Artificial Neural Networks (ANN) that joins Support Vector Machine (SVM) in a symmetric and asymmetric ways. This approach can deal with the classification inhomogeneity between various classifiers. In addition, this approach can deal with the imbalance problems in the CBIR algorithm. This method has been created because the retrieval results in multi-class search environment are so far from the expected. That is for the overlapping semantics between different classes. The experiments carried out on the Corel, Caltech-101 and COIL DBs, and proved the effectiveness of the proposed method compared to the state-of-arts methods.

The usage of the robust and consistent ETL stages in DW was better in achieving high efficiency. That is by

enhancing the preprocessing tasks. As well as, diminishing noise rates and using efficient CBIR that was better for getting top-ranked images. That is by reducing the possibilities of results and that pours into accuracy. Therefore, I will use the “Content-based” that implies the search analysis of the contents of the image. Using this approach instead of the metadata such as keywords, labels, or descriptions related to the image. The term “content” in this setting might allude to colors, shapes, textures, or any other data that can be determined from the image itself. CBIR algorithm is alluring since looks that depend simply on metadata are subordinate to content quality and completeness.

### B. Image Similarity-Based Measures

The author in [3], the Memetic algorithm is conducted to retrieve the top-ranked that is relevant to the entered query. The MA-based similarity measures do many operations. These operations are to extract an effective color, a shape and color texture features from the Corel image database. The ILS (Iterated local search) algorithm assesses the quality of the solution by raising the fitness function number. This method succeeds in increasing the average of precision and recall by 1.4 and 8.05 respectively.

One of the most achieving works is [3] and chose it to be the guide in the research. One of the most well-known metaheuristics in dealing with optimization issues could be GA. GA may be a population-based heuristic approach. GA bargains with a major framework to discover an appropriate arrangement for an uneasy issue. In addition, it has demonstrated its proficiency in producing unused arrangements between trial arrangements and recognizing the finest solutions from a populace. It is a must to have a nearby look to survey GA in finding the arrangement space. That was rather than caring approximately abusing the search space that is for improving the metaheuristic algorithm.

There are inexhaustible computer imaging applications requiring a few kinds of similarity estimation as the portion of their forms. Although the applications are much shifted and the execution subtle elements of each arrangement are special, all share the common string in those features or properties of the image (in each application). These features are measured and after that compared to other features from a database of images. In addition, these features would be compared to a few references show to extricate a few significant conclusions or usefulness approximately the image information on hand. There are a few strategies for measuring image similarity. The strategies are: 1) A pattern recognition approach. 2) Comparison of outlines in a video sequence. 3) Image stabilization employing a homographic transformation. 4) Utilizing image feature focuses to compute similarities and generate an image mosaic. Here, this paper will deal with the fourth strategy of measuring similarity search as the previous research. This paper will use this strategy because it will produce more accurate results regarded to the research of that section.

### C. Image De-noising

The author in [13], Wei Wang examined a filtering method called novel inverse by employing a dictionary approach. The most thought is to combine a learned word reference for the representation of the de-convoluted image. As well as, a reverse channel based on non-negativity and bolster limitations. That is to de-convoluted the watched image with an obscure point spread function. The advantage of this approach is that the objective image can be spoken to with more points of interest by the word reference premise. This paper utilizes the substituting course strategy of multipliers (ADMM) to unravel the coming about optimization issue. The limitations are the inability to recognize the blurring data region and the support region. In this work, an image-preparing model is created, whereas an image-filtering module is created using the median channel with a variable versatile window design. This filtering provides change inters of working recurrence and PSNR. The moment stage includes the advancement of a compression module based on lifting wavelets. This compression module provides improved execution inters of compression proportion and PSNR. In the final stage, image encryption is performed to have a secured image [1]. The drawback does not use different filtering techniques to choose the best results.

Here, in [14] a new strategy presents a gradient-based approach for extracting the notable structures from diverse multimodal medical images. That is for the estimation of combination execution and consequent determination. The benefits of this strategy are showing the leading performance of visual perception joint quantitative metrics and the weakly involved calculations. The only limited point that observed here is the inactive strategy with big data. An unused de-noising approach is proposed for rebuilding the Gaussian (AWGN) noise-contaminated images. In addition, through the suggested developmentally optimized adjust Guided Image Filtering [15]. The short-come of this approach is dealing with only one type of noise (Gaussian noise). This research introduces the infrared and visible image fusion algorithm relayed on the image division of a hybrid curvature filter [16]. The limitations of this algorithm are low performance and low enhancement of the running speed.

In this research, evolving an Image Filtering and Labeling Assistant (IFLA) framework is to assist the foremost time-consuming parcel of this handle [17]. The drawbacks of this research are the inability to automatically label the marked-up ranges of images. Moreover, the inability to meet behavioral patterns and programmed individual symmetry. This paper portrays Gaussian filter and Histogram Equalization methods for de-noising and differentiates the upgrade of MR Brain images separately [18]. The suggested future work for this paper, is removing the other types of noise by using the comparing filter.

Another paper, innovates an unused filter for image boosting. This filter relayed on the combination of a shock channel with a fourth-order dissemination equation [19]. The limited result was noticed while applying this algorithm with images of degraded documents. The new

algorithm combines the versatile median filter algorithm with the conventional nonlocal mean algorithm. To begin with that, you should alter the image window and adaptively choose the comparing pixel weight. Then de-noises the image, which can have a great filtering impact on the blended noise [20]. The drawbacks of this calculation are discovering the calculation for color image processing effect and enhancing the time consumption of de-noising image.

The author in [21], point-by-point comparative investigation of diverse de-noising filtering techniques. This technique implemented leaned on four widely used Image Quality Assessments (IQA) metrics. The drawback of this paper is not diminishing more types of image noise. Lufeng Bai, et al. [22] designed a singular method for Cauchy method removal that illuminates the new demonstration. Then a rotating minimization strategy is utilized, and its merging is demonstrated. The limitation of this method is not calculating the precision and recall values to evaluate the system's efficiency.

The proposed strategy [23] is organized utilizing three modules, to be specific. These strategies are: 1) The DIP-based module for concurrently learning noise diminishment and differentiate upgrade. 2) On the off chance that module for combining images and neutralizing color movement and noise enhancement. 3) The PR module for enhancing edge data. The drawback of this strategy is the inputs of the image fusion module can delay the velocity of the whole model. A novel variational paradigm [24] is presented based on adding up to a variety. As well as IO the gauge for at the same time expelling the tangling noise, assessing the location of lost pixels, and filling in them. The suggested issue for enhancing the quality is using a similarity search algorithm.

All of the previous papers proposed various kinds of de-noising algorithms and strategies. Most of these used Gaussian filter lonely or a hybrid of Gaussian filter with different methods for de-noising and upgrading the performance of the targeted image. In addition, it was notable that the medical image has a dominant hand in that issue. Most of the shortcomings were from the use of one type of filtering technique in addition to low performance and running speed.

The proposed paper will introduce a new algorithm CEDF (Color Edge Detection with Gaussian Blur Filter). CEDF applies the Gaussian Blur Filter after using the CED method for feature detection of the image. Furthermore, to assess the efficiency of the system's result, will figure precision and recall values, and show the results applied to Correl DB. By contribution of the Gaussian Blur filter with the proposed method, the results were applied to Correl DB. Moreover, rising through recall and precision average 1.4 and 8.05 respectively and assess the accuracy of the result.

#### IV. PROPOSED ALGORITHM (CEDF)

Multimedia data play an essential role in the decision process. One of the problems when integrating multimedia

data in a warehouse is the retrieval speed of the information. This information is a direct coefficient with respect to efficiency and dealing with dimensions built on information retrieval [10]. The use of DW is most common for storing/ retrieving complex data in multimedia formats. Where DBMS algorithm is traditional and is not designed to fit in handling complex multimedia data, since relational databases store structured data. That is why enhancing warehouse performance on multimedia data is essential. Because of that, this paper deals with boosting the performance of DW. That is activated by upgrading the efficiency of the algorithm with the help of promoting precision and recall results [3].

The proposed method will show how to extract features from a set of images (color, shape and texture features) by using the CBIR algorithm with Color Edge Detection (CED) method. Once these features are extracted, the method will minimize the distance between these features vectors and the query image one, using a genetic algorithm. This paper illustrates the extraction of endless strong and imperative features from the database of images. Moreover, the capacity of these features is within the store through the frame of features vectors. Accordingly, an imaginative closeness assessment with a metaheuristic algorithm (GA with SA) has been attained. This attainment is between the QI features and those having a place in the database image. Finally, the paper will measure the precision and recall of this method. By contribution of the Gaussian Blur filter with the proposed method, the results applied to Correl DB. As mentioned before that the extraction of the image features depended on three basic phases (color, shape, and texture). The paper will use the second and the third phases as it used in the previous paper [3]. The enhancement of the paper only includes the color phase by using (CED) algorithm that we will enumerate in details in the following section.

##### A. Color Edge Detection

Edge detection is the title for a set of scientific strategies, which target classifying focuses in an image. That is because the image concentrated changes strongly or, has discontinuities. Edge detection alludes to the method of distinguishing and finding sharp discontinuities in an image. Edge detection could be a principal device utilized in most image-preparing applications. That is to get data from the outlines as an antecedent step to feature extraction and object segmentation [25]. One of the elemental errands in image-preparing is edge detection. High-level image handlings, such as object acknowledgment, division, image coding, and robot vision depends on the precision of edge detection. Whereas edges are contain fundamental image data.

Within the conventional color edge detection techniques, the color image is, to begin with, isolated into three distinctive channels. These channels are R, G, and B before handling and after that, the algorithm is connected to the particular channels. Additionally, in conclusion, all channels are combined to deliver the result. The high-level image handling applications such as question acknowledgment, protest following, robot vision, etc. In the

case of color image processing, Color images require more memory space for capacity. Unlike the greyscale, images require the transmission of color data requires a bigger transfer speed. By utilizing an effective edge discovery procedure, the unnecessary subtle elements of a color image can be disposed of. In addition, the valuable data can be put away to assist the preparation. This will successfully diminish the memory space for putting away the color data and lower the transmission bandwidth [26]. The edge discovery methods can be broadly classified as:

- Edge detection in grey-scale images.
- Edge detection in color images.

The basic distinction between a gray-scale image and a color image is the pixel in a gray-scale image may be a scalar esteemed work. While in a color image, a pixel is considered as a vector esteemed work because it comprises three-color components (red, green and blue). Due to this, vector esteemed procedures are favored for edge detection in color images [26].

1) *An outline of a few classic edge detectors:* Many edge detection methods have been proposed for grey-scale images in the image handling literature. These edge detectors are categorized as [26]:

- Edge detection strategies are based on the first-order subordinate.
- Edge discovery strategies are based on the second-order subsidiaries.

#### 2) *The four steps of Edge Detection*

- Filtering- Filter image to move forward execution of the edge detector w.r.t noise error.
- Smoothing- stifle as much noise as conceivable, without wrecking the true edges.
- Enhancement- applies a channel to improve the quality of the edges within the image (sharpening).
- Detection- decides which edge pixels ought to be disposed of as noise and which ought to be held (as a rule, thresholding gives the basis used for detection). Localization- decide the precise location of an edge (sub-pixel resolution may well be required for a few applications. That appraise the area of an edge to better than the dispersing between pixels). Edge diminishing and connecting are as a rule required in this step. Edge discovery could be a principal apparatus utilized in most image-handling applications to get data from the outlines. Furthermore, a forerunner step includes extraction and protest segmentation [25].

#### B. *How Does CEDF Works*

Content-Based Information retrieval (CBIR) algorithm features are extracted through several techniques for the features extraction. Most visual images contain color and shape, with low-level image features and notable focuses [3]. Above all, of that, each image stores its attributes in a partitioned database called database features. A similar

CBIR algorithm is outlined in Fig. 1. Here [3], is the recent paper that talked about CBIR and will reuse some steps and add another step to my new paper. On one hand the steps that will use are: (1) how to extract features from a set of images (color, shape, and texture features). (2) Once these features are extracted. (3) minimize the distance between these features vectors and the query image one using GA.

Finally, will measure the precision and recall of this method and show the results applied to Correl DB. On the other hand, we will add a new filtering technique called Gaussian Blur filtering technique with the old method. This technique concludes (RGB color with neutrosophic clustering calculation, Canny Edge strategy to extricate shape features, YCbCr color with discrete wavelet change, Canny Edge Histogram to extricate color features, and gray-level co-occurrence framework to extricate texture features). This filtering technique will raise the average precision and recall percentage and that reflects positive performance results on the proposed algorithm.

1) *CEDF algorithm steps:* Fig. 1 demonstrates the method suggested. The fundamental steps involved in color feature extraction are shown as follows:

- Select the image query to be processed.
- Image query and DB images will be transformed into YCbCr co; or space from the M\*N input RGB image.
- Before applying the canny edge detector to the Y portion of the image, the portion of YCbCr is separated following the transformation of the image.
- After that, the unobtrusive Cb and Cr are combined with the edge that was obtained in the previous step.
- Pixel editing and the addition of CED with a noise-eliminating filtering method.
- After that, a single RGB image is created from the synthesized image.
- The histogram calculation for each portion of the RGB image is followed by the division of the RGB component into distinct R, G, and B components. From HR, HG, and HB, a total of 256 bins have been obtained.
- In each histogram that is obtained, a distinct wavelet transform is used to improve the performance of the features. to apply DWT level 3 to HG and HB and DWT level 2 to HR. There are 128 bins created when DWT is applied. In particular, 64 bins were obtained from HR, while each HG and HB obtained 32 bins.
- The repository is where each image's feature vector is calculated.
- Similarity search algorithms will elaborate to get the most similar images.
- Get the results.

The following section will illustrate the main features of (CBIR) algorithm features.

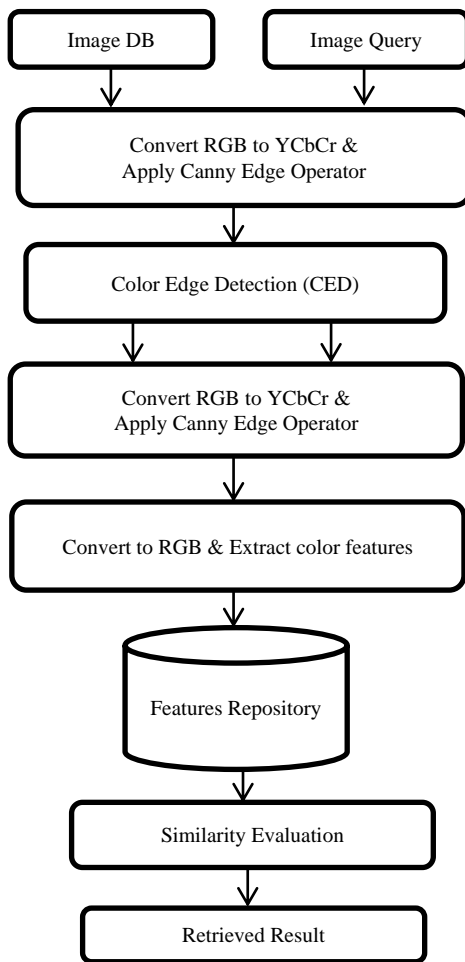


Fig. 1. The proposed method for color feature extraction.

2) *Color features extraction*: In this modern work, the strategy for CBIR algorithm based on the extraction of the color features, are mentioned. This strategy [3] is concluding the conversion of RGB to YCbCr and applies the canny edge detector to the result. Then will use the CED operation that is concluding the proposed filtering technique. The next sections will present the details of how to convert RGB to YCbCr, using the canny edge detector and what are the steps of CED operation.

Color is the foremost broadly utilized visual substance for image recovery. Its three-dimensional values make its separation possibility predominant to the single-dimensional gray values of images. Sometimes recently selecting a suitable color depiction, color space must be decided to begin with. Each pixel of the image can be spoken to as a point in a 3D color space.

a) *Query image conversion and CED*: Commonly utilized color space for image recovery incorporates RGB, Munsell, and CEDF. In the proposed paper case, the method will choose RGB and Y'CbCr spaces. Y'CbCr color spaces are defined by a mathematical coordinate transformation from an associated RGB color space. If the underlying RGB color space is absolute, the Y'CbCr color space is an absolute color

space as well. Conversely, if the RGB space is ill-defined, so is Y'CbCr. The formula that allows converting RGB colors to Y'CbCr is the following:

$$[Y \ Cb \ Cr] = [RGB] \begin{Bmatrix} 0.229 & -0.168935 & 0.499813 \\ 0.587 & -0.331665 & -0.418531 \\ 0.114 & 0.50059 & -0.081282 \end{Bmatrix} \quad (1)$$

After converting the RGB to YCbCr, one should apply the canny edge detector to the result [3]. Edges in any image are helpful in identifying the objects. The new step that is added to the old method is Color Edge Detection (CED). The CED works by following these steps:

- Apply Gaussian Filter to smooth the image to remove the noise.
- Find the intensity gradients of the image.
- Apply non-maximum suppression to get rid of spurious responses to edge detection.
- Apply a double threshold to determine potential edges.
- Track edge by hysteresis: Finalize the detection of edges by suppressing all the other edges that are weak and not connected to strong edges.

b) *Conversion to RGB*: After applying CED, the method reconverts the image to RGB and computes the histogram for every component (R, G and B). The color histogram serves as a viable representation of the color substance of an image in the event. That color design is one of a kind compared with the rest of the DB. The color histogram is simple to compute and successful in characterizing both the worldwide and nearby dispersions of colors in an image. In expansion, it is strong to interpretation and turns almost the seeing pivot. Moreover, the changes as it were gradual with the scale, impediment, and seeing angles.

3) *Shape features*: The shape of an object is an important and basic visual feature that can describe image content. In the context of content-based image retrieval, the word shape is used to refer to the geometry of a region's bounding contour in 2D. Shape feature extraction and representation are the bases of object recognition in an image. It plays an important role in many image-processing applications including content-based image retrieval. The feature extraction stage produces a representation of the content that is useful for shape matching. The easiest way to use geometric information about the objects present in an image is by computing simple scalar descriptors of the regions of interest, such as area, circularity, eccentricity, major axis orientation, dimensions of the minimum bounding rectangle, etc. Usually, the shape representation is kept as compact as possible for the purpose of efficient storage and retrieval, and it integrates perceptual features that allow the human brain to discriminate between shapes.

Finally, here are the steps followed before computing the invariant moments:

- Convert the image 2D matrix to a gray matrix.



- Apply the median blur filter on the gray matrix.
- Apply the Canny Edge detector on the filtered image.
- Generate the moments and then the Invariant moments of the image.
- Store the invariant moments in the features vector.
- Convert the image 2D matrix to a gray matrix.
- Apply the median blur filter on the gray matrix.
- Apply the Canny Edge detector on the filtered image.
- Generate the moments and then the Invariant moments of the image.
- Store the invariant moments in the features vector.

4) *Texture features*: Texture feature is a vital property of the image, which is often characterized as the visual appearance, or material characteristics of the objects within the image. It basically comprises the components of surface primitives (i.e., text components or textual) organized in a few indicated arrange (i.e., text format). The idea of a textual is central to the text. For the most part, it is characterized as a visual primitive that shows certain invariant properties more than once very different positions, distortion, and introduction in a given zone. The text components can be as little as a sand molecule to huge components like bricks in dividers, but they all share the comparative auxiliary and measurable properties inside a gather. The textual properties deliver rise to the seen delicacy, consistency, thickness, harshness, consistency, linearity, recurrence, stage, directionality, coarseness, arbitrariness, fineness, smoothness, granulation, etc., of the text as an entirety.

## V. PERFORMANCE EVALUATION

### A. Database Descriptions

The tests conducted in this consider utilized the Corel DB (A) comprising 1000 differing images and each image within the DB is  $384 * 256$  or  $256 * 384$  size-wise. As such, the results were expressed with the utilization of 10 semantic sets. As well, each set contains one hundred pictures. The Corel DB is of the taking after semantic categories. These categories are Mountains, Africa, Dinosaurs, Buildings, Buses, Nourishment, Elephants, Horses, Beaches, and Flowers. Consequently, with the utilization of the recorded comes about, a clear comparison of comes about can be performed.

Then the proposed paper will measure the precision and recall percentage for every category in the DB by the proposed method (CEDF). After that, the paper will compare the results with some other methods related to the topic of the paper's interest.

For that will appear the adequacy and prevalence of the proposed method in terms of recovery capabilities. Those capabilities will show by the gotten results are too compared with other algorithms in terms of precision and recall rates.

The other comparative frameworks are J. C. Felipe, Z. Mehmood, T. Mahmood, and M. A. Javid [27], N. Ali, K. B. Bajwa, R. Sablatnig, and Z. Mehmood [28], Z. Mehmood, S. M. Anwar, N. Ali, H. A. Habib, and M. Rashid [28]. Zeng, R. Huang, H. Wang, and Z. Kang [29], E. Walia and A. Pal [30], C. Wang, B. Zhang, Z. Qin, and J. Xiong [3], M. K. Alsmadi [3] and the same ten semantic sets (each set contains 100 images of Corel DB) were utilized by these comparative algorithms. Tests of diverse images from 10 semantic sets of the Corel DB appear in Fig. 1.

### B. Precision and Recall Evaluation

To assess the execution and the viability of the proposed CED framework, precision and recall rates are used. Precision could be a degree of the capacity of the CED algorithm to recover. So those images will be compared to the query image. In this respect, the recall rate is the genuine positive rate, and it is utilized to assess the ability of the CED framework. That ability for regarding the number of recovered important images compared to all similar images within the database.

The precision of a CED algorithm is defined as the number of relevant images retrieved as a fraction of the total number of images retrieved [3]:

$$Precision = \frac{\text{number of relevant images retrieved}}{\text{total number of images retrieved}} \quad (2)$$

The value of Precision lies between 0% and 100%.

Another measure to find the effectiveness of a retrieval method is Recall. Recall of a CED algorithm is defined as the number of relevant images retrieved as a fraction of the total number of relevant documents in the database:

$$Recall = \frac{\text{number of relevant images retrieved}}{\text{total number of relevant images in the database}} \quad (3)$$

Recall values have range similar to precision, e.g., between 0% and 100%.

The recall value will be 100% if the entire database is returned as a result set in response to a query. Similarly, the precision value can be kept high by retrieving a small number of images. Therefore, precision and recall should be used together or the number of images retrieved should be specified [3].

The results of the execution of the paper's test were stated by running the proposed CED algorithm simulation five times. In every experiment, 20 inquiry images from each semantic category have been chosen arbitrarily. Then the rates of precision and recall are expressed based on the most elevated 20 images. As can be observed in Fig. 2 the exactness and review rates were calculated for 20 queries in the expansion to their recovered images. As Fig. 2 is illustrating the proposed algorithm contains a great level of effectiveness for the retrieval of images. These images were retrieved by getting higher 5 categories of precision and recall values. As the tests were demonstrating, the increment within the retrieved comparative images will move forward with precision and recall rates. The results on the extracted color, shape, and texture features

combined with the MA component. That poured in demonstrating the exceptionally empowering enhancements of the CBIR algorithm, especially regarding its recall and precision.

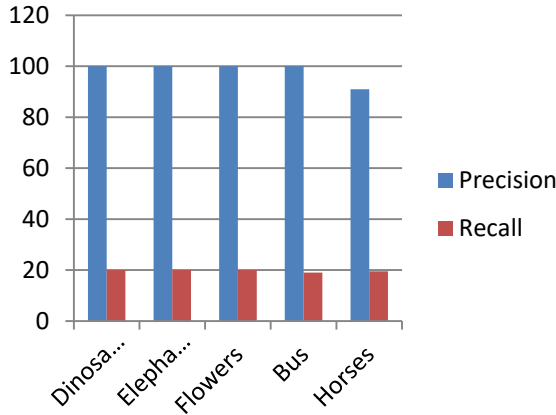


Fig. 2. Precision and recall rates of CED [3]

### C. CBIR with Filters CED with /without Filters

To improve the results, should apply some filters to the original image before applying the CED algorithm. Here are the filters descriptions:

- **Blur:** Image blurring is achieved by convolving the image with a low-pass filter kernel. It is useful for removing noise. It removes high-frequency content (e.g. noise, edges) from the image. So, edges are blurred a little bit in this operation (there are also blurring techniques that do not blur the edges).
- **Gaussian Blur:** Blurs an image using a Gaussian filter.
- **Median Blur:** While the “Gaussian” blur filter calculates the mean of the neighboring pixels, the “Median” blur filter calculates the median.
- **Laplacian:** Laplacian filters are derivative filters used to find areas of rapid change (edges) in images. Since derivative filters are very sensitive to noise, it is common to smooth the image (e.g., using a Gaussian filter) before applying the Laplacian. This two-step process is called the Laplacian of Gaussian operation.

The paper used the Corel database for testing purposes. It runs the CED on 20 random query images for each class using the formula proposed in “Precision and Recall” section (Eq. 2, Eq. 3). Those will trial without adding any filtering techniques and with filtering techniques. Without filters, the recall values for the proposed (CED) method are below the average of the proposed (CED) method. That is when the Gaussian filter comparative to the other filters. For that reason, the paper will test some filters that were applied to the query image and the DB before comparing. Fig. 3 shows the recall values for some sets in Corel DB with and without filters. In addition, the results confirm that the Gaussian blur filter has better recall values than using the other different filters or without using any filters.

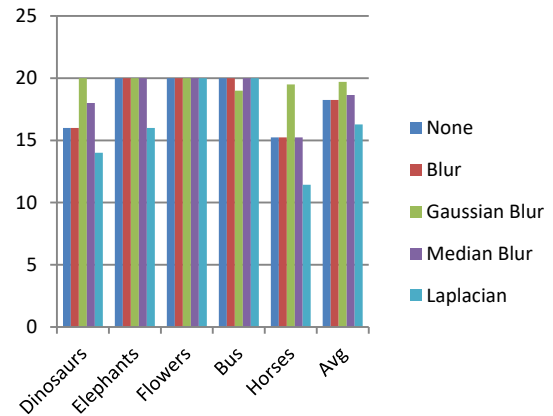


Fig. 3. Comparisons between using the CED method with and without filters (recall measure).

On one hand, Fig. 3 shows that the Gaussian Blur filter gives better results for all random proposed classes of the DB. On the other hand, the Laplacian filter gives the worst results compared to the others. Consequently, use the Gaussian blur filter with CED method to get better performance and more accurate results.

Without filters, the precision values for the proposed (CED) method are below the average of the proposed (CED) method. That is the case of the Gaussian filter compared to the other filters. For that reason, the paper tested some filters that were applied to the query image and the DB before comparing. Fig. 3 shows the precision values for some sets in Corel DB with and without filters. In addition, the results confirm that the Gaussian blur filter has better precision values than using the other different filters or without using any filters.

On one hand, Fig. 4 shows that the Gaussian Blur filter gives better results for all random proposed classes of the DB. On the other hand, the Laplacian filter gives the worst results compared to the others. Therefore, the paper use the Gaussian blur filter with CED method to get better performance and results that are more accurate.

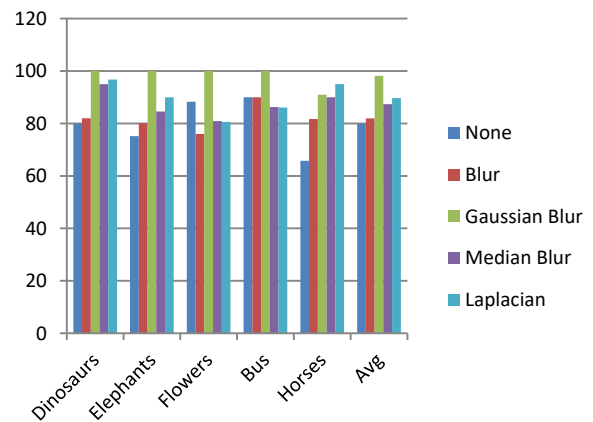


Fig. 4. Comparisons between using the CED method with and without filters (precision measure).

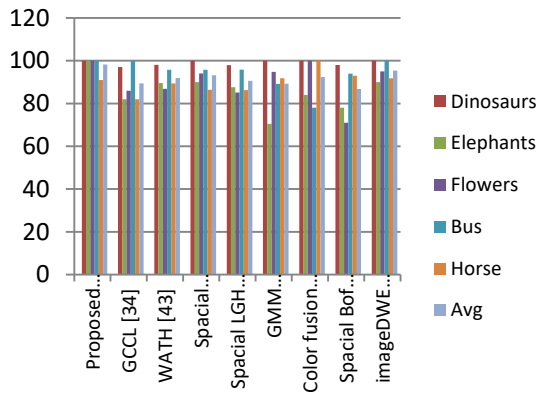


Fig. 5. Comparisons between the proposed method (CEDF) and other state-of-the-art algorithms (recall measure).

This figure shows that CEDF method gets the highest average over all the other methods. That clears especially on dinosaurs, elephants and flowers. That average is 18.31565% on the recall measure and that proves the effectiveness of the proposed method. That is when the paper compares its results to the results in Fig. 2 that proves the importance of penetrating a new filter (Gaussian filter) to the proposed method CEDF. It will be obvious from these figures (Fig. 3 & 5) that the proposed algorithm has better average recall over different points while using various methods for Corel DB.

Fig. 6 shows that the CEDF method gets the highest average over all the other methods. That average is especially on dinosaurs, elephants and flowers by scoring 100% on the precision measure. In addition, that proves the effectiveness of the proposed method. That is when the paper compares its results to the results in Fig. 2 that run the CED on 20 random query images for each class using the formula proposed in “Precision and Recall” section. This formula is without adding any filtering techniques and that proves the importance of penetrating a new filter (Gaussian filter) to the proposed method CED.

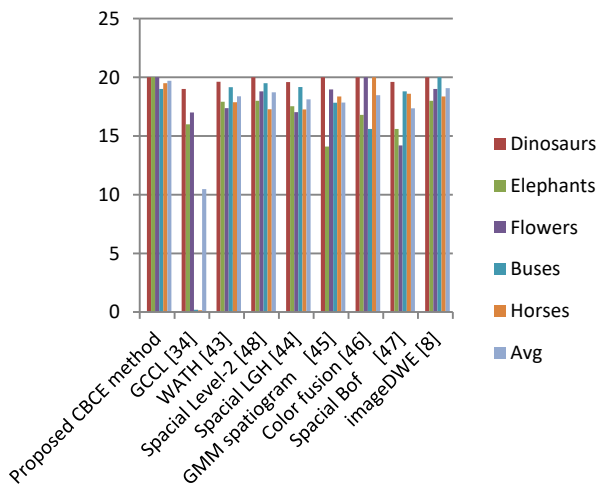


Fig. 6. Comparisons between the proposed method (CEDF) and other state-of-the-art (precision measure).

## VI. EVALUATION AND DISCUSSION

In this section, the paper depicts three experiments dealing with diverse images represented by diverse descriptors. For each context with and without using the filtering technique, took similarity rates to be compared. This comparison will be with each other in the two cases of using the recall and precision values.

Experiment 1: This experiment is based on examining the recall and precision rates when the paper is using different filtering techniques without using any filtering technique. The result of that experiment shows the high rates of recall and precision results. That is when the paper is using the filtering technique with the same DB and that proves the effectiveness of using a filtering technique (Fig. 3 and Fig. 4).

Experiment 2: (recall measure). This experiment depended on comparing the results of the proposed method CEDF. That experiment contains a filtering technique with the other methods of different researchers. Those researchers use the same DB but without using any filtering technique. The results of that experiment show the increasing rate of recall values by 1.4 and that proves the effectiveness of the proposed method CEDF Fig. 5.

Experiment 3: (precision measure). This experiment depended on comparing the results of the proposed method CEDF. That experiment contains a filtering technique with the other methods of different researchers. Those researchers use the same DB but without using any filtering technique. The results of that experiment show the rising rate of precision values by 8.05 and that proves the effectiveness of the proposed method CEDF Fig. 6.

The average precision and recall values of the proposed CEDF method fared better than the other cutting-edge methods, by raising the values by 8.05 and 1.4, respectively.

This is due to the fact that the other developed methods only extract a small number of features from various descriptors including shape, color, and texture. While in this work, measurements of color, shape, and texture were used to extract a large and robust number of characteristics.

Specifically, color characteristics were extracted using YCbCr color with discrete wavelet transform and Canny edge histogram, shape features were extracted using RGB color with neutrosophic clustering algorithm and Canny edge technique, and texture features were extracted using GLCM.

By raising the fitness number, the SA algorithm's combined use with GA increased the solution quality (weight). Because of this, adding the SA algorithm improved the GA's exploitation rather than its exploration of the search space. The experimental findings make it abundantly evident that the MA aids in the retrieval of a significant number of photos that are pertinent to the query image [3].

## VII. CONCLUSION AND FUTURE WORKS

This study is advancement in querying unstructured medical data using a similarity search algorithm. Specifically, the contribution of this work is providing unlimited query support and efficiency in processing unstructured data in the epilepsy field. At the core of the queries are modules, which perform feature extraction from unstructured data. The framework is unlimited by adding CEDF, which has a filtering technique to erase noise and editing pixels. This filtering is done for the query image in the feature extraction phase. The framework is efficient in that it utilizes the distributed computing power of Gaussian filters.

The study shows how to extract features from a set of images (color, shape and texture features). Once these features are extracted, the minimization of the distance between these features vectors and the query image one using a genetic algorithm. Finally, we will measure the precision and recall of this method and show the results applied to Correl DB. By contribution of the Gaussian Blur filter with the proposed method, the results rising through recall and precision average 1.4 and 8.05, respectively and that assesses the accuracy of the result. As a result, such a framework is feasible and useful for medical queries. Although this study is specific to the epilepsy field, however, it is a step forward in data-driven medicine. The data-driven medicine content of the unstructured data is available for unlimited and efficient querying.

After the paper produces a concrete DW infrastructure, the next research will recommend looking into integration and usability issues.

## ACKNOWLEDGMENT

I must first express my gratitude towards Prof. Ahmed Sharaf Eldin for her supervision and for teaching me how to be an independent thinker, she has been an invaluable source of knowledge and ideas and I appreciate her flexibility in allowing me to pursue my own research ideas.

My sincere thanks are dedicated to Dr. Nermin Abdelhakim for his generous assistance and continuous help in addition to providing me with all the facilities required for the success of this work.

## REFERENCES

- [1] S. K. Shelke, S. K. Sinha, and G. S. Patel, "Development of complete image processing system including image filtering, image compression & image security," *Mater. Today Proc.*, no. xxxx, 2021, doi: 10.1016/j.matpr.2021.06.154.
- [2] G. Blaži, P. Poš, and D. Jakši, "Data Warehouse Architecture Classification," pp. 1491–1495, 2017.
- [3] M. K. Alsmadi, "Content-Based Image Retrieval Using Color, Shape and Texture Descriptors and Features," *Arab. J. Sci. Eng.*, vol. 45, no. 4, pp. 3317–3330, 2020, doi: 10.1007/s13369-020-04384-y.
- [4] R. M. Hawlick, "International Conference on Computer Vision and Image Processing, CVIP 2016," *Adv. Intell. Syst. Comput.*, vol. 459 AISC, no. 5, pp. 1–644, 2017.
- [5] D. R. Murphy et al., "Computerized Triggers of Big Data to Detect Delays in Follow-up of Chest Imaging Results," *Chest*, vol. 150, no. 3, pp. 613–620, 2016, doi: 10.1016/j.chest.2016.05.001.
- [6] A. Nikonorov, S. Bibikov, V. Myasnikov, Y. Yuzifovich, and V. Fursov, "Correcting color and hyperspectral images with identification of distortion model," *Pattern Recognit. Lett.*, vol. 83, pp. 178–187, 2016, doi: 10.1016/j.patrec.2016.06.027.
- [7] S. Unar, X. Wang, C. Wang, and Y. Wang, "A decisive content based image retrieval approach for feature fusion in visual and textual images," *Knowledge-Based Syst.*, vol. 179, pp. 8–20, 2019, doi: 10.1016/j.knosys.2019.05.001.
- [8] L. Ma, X. Liu, Y. Gao, Y. Zhao, X. Zhao, and C. Zhou, "A new method of content based medical image retrieval and its applications to CT imaging sign retrieval," *J. Biomed. Inform.*, vol. 66, pp. 148–158, 2017, doi: 10.1016/j.jbi.2017.01.002.
- [9] K. T. Ahmed, S. Ummesafi, and A. Iqbal, "Content based image retrieval using image features information fusion," *Inf. Fusion*, vol. 51, no. December 2017, pp. 76–99, 2019, doi: 10.1016/j.inffus.2018.11.004.
- [10] K. N. Sukhia, M. M. Riaz, A. Ghafoor, S. S. Ali, and N. Iltaf, "Content-based histopathological image retrieval using multi-scale and multichannel decoder based LTP," *Biomed. Signal Process. Control*, vol. 54, p. 101582, 2019, doi: 10.1016/j.bspc.2019.101582.
- [11] K. V. Madhavi, R. Tamilkodi, and K. J. Sudha, "An Innovative Method for Retrieving Relevant Images by Getting the Top-ranked Images First Using Interactive Genetic Algorithm," *Procedia Comput. Sci.*, vol. 79, pp. 254–261, 2016, doi: 10.1016/j.procs.2016.03.033.
- [12] A. Irtaza et al., "An ensemble based evolutionary approach to the class imbalance problem with applications in CBIR," *Appl. Sci.*, vol. 8, no. 4, pp. 1–26, 2018, doi: 10.3390/app8040495.
- [13] W. Wang and M. K. Ng, "Dictionary-based inverse filtering methods for blind image," *Appl. Math. Model.*, vol. 96, pp. 269–283, 2021, doi: 10.1016/j.apm.2021.03.005.
- [14] B. Goyal, D. Chyophel, A. Dogra, and V. Bhateja, "Measurement and analysis of multi-modal image fusion metrics based on structure awareness using domain transform filtering," vol. 182, no. June, 2021, doi: 10.1016/j.measurement.2021.109663.
- [15] H. Singh, S. V. R. Kommuri, A. Kumar, and V. Bajaj, "A new technique for guided filter based image denoising using modified cuckoo search optimization," *Expert Syst. Appl.*, vol. 176, no. March, p. 114884, 2021, doi: 10.1016/j.eswa.2021.114884.
- [16] G. Liu et al., "composition," *Infrared Phys. Technol.*, p. 103938, 2021, doi: 10.1016/j.infrared.2021.103938.
- [17] T. Xi, J. Wang, H. Qiao, C. Lin, and L. Ji, "Ecological Informatics Image Filtering and Labelling Assistant (IFLA): Expediting the analysis of data obtained from camera traps," *Ecol. Inform.*, vol. 64, no. June, p. 101355, 2021, doi: 10.1016/j.ecoinf.2021.101355.
- [18] K. S. Reddy and T. Jaya, "Materials Today : Proceedings De-noising and enhancement of MRI medical images using Gaussian filter and histogram equalization," *Mater. Today Proc.*, no. xxxx, 2021, doi: 10.1016/j.matpr.2021.03.144.
- [19] S. Thierry, W. Colince, N. Eloundou, and N. Alexandre, "Shock filter coupled with a high-order PDE for additive noise removal and image quality enhancement," *Array*, vol. 12, no. June, p. 100105, 2021, doi: 10.1016/j.array.2021.100105.
- [20] K. Huang and H. Zhu, "Image Noise Removal Method Based on Improved Nonlocal," vol. 2021, 2021.
- [21] S. M. Boby, "Medical Image Denoising Techniques against Hazardous Noises : An IQA Metrics Based Comparative Analysis," no. April, pp. 25–43, 2021, doi: 10.5815/ijigsp.2021.02.03.
- [22] L. Bai, "A new approach for Cauchy noise removal," vol. 6, no. April, pp. 10296–10312, 2021.
- [23] S. Huang, Q. Hoang, T. Le, Y. Peng, and C. Huang, "An Advanced Noise Reduction and Edge Enhancement Algorithm," pp. 1–12, 2021.
- [24] M. A. Khan, F. A. Dharejo, F. Deeba, S. Ashraf, J. Kim, and H. Kim, "Toward developing tangling noise removal and blind inpainting mechanism based on total variation in image processing blind inpainting mechanism based on total variation in image processing," no. March, pp. 2–5, 2021, doi: 10.1049/ell2.12148.
- [25] V. Banga, "Improved Color Edge Detection by Fusion of Hue , PCA & Hybrid Canny," vol. 2, no. 1, pp. 1–8, 2015.

- [26] S. Subhasini, "Color Image Edge Detection: A Survey," *Int. J. Innov. Eng. Technol.*, vol. 8, no. 1, pp. 235–247, 2017, doi: 10.21172/ijiet.81.032.
- [27] Z. Mehmood, T. Mahmood, and M. A. Javid, "Content-based image retrieval and semantic automatic image annotation based on the weighted average of triangular histograms using support vector machine," *Appl. Intell.*, vol. 48, no. 1, pp. 166–181, 2018, doi: 10.1007/s10489-017-0957-5.
- [28] Z. Mehmood, S. M. Anwar, N. Ali, H. A. Habib, and M. Rashid, "A Novel Image Retrieval Based on a Combination of Local and Global Histograms of Visual Words," *Math. Probl. Eng.*, vol. 2016, 2016, doi: 10.1155/2016/8217250.
- [29] S. Zeng, R. Huang, H. Wang, and Z. Kang, "Image retrieval using spatiograms of colors quantized by Gaussian Mixture Models," *Neurocomputing*, vol. 171, pp. 673–684, 2016, doi: 10.1016/j.neucom.2015.07.008.
- [30] E. Walia and A. Pal, "Fusion framework for effective color image retrieval," *J. Vis. Commun. Image Represent*, vol. 25, no. 6, pp. 1335–1348, 2014, doi: 10.1016/j.jvcir.2014.05.005.

# Agro-Food Supply Chain Traceability using Blockchain and IPFS

Subashini Babu, Hemavathi Devarajan

Department of Data Science and Business Systems-School of Computing, SRM Institute of Science and Technology  
Chengalpattu, Kattankulathur, Tamilnadu 603 203, India

**Abstract**—Many blockchain initiatives significantly use the InterPlanetary File System (IPFS) to store user data off-chain. The centralized administration, ambiguous data, unreliable data, and ease of creating information islands are all issues with traditional traceability systems. This study develops a monitoring system using blockchain technology to record and inquire about product information in the supply network of Non-Perishable (NP) agro goods to address the above issues. The transparency and trustworthiness of traceability data were considerably improved by employing blockchain technology's distributed, tamper-proof, and traceable properties. To alleviate the strain on the blockchain and enable efficient information inquiry, a storage structure is built in which both public and private data are stored in the blockchain and the Inter Planetary File System (IPFS) using cryptography. Because of its ability to trace the origin of food, blockchain technology contributes to the development of reliable food supply chains and the establishment of rapport between farmers and their customers. Since it provides a secure location for data to be kept, it can pave the way for implementing data-driven farming techniques. In addition to improving data security, recording farm data in IPFS and storing encrypted file IPFS hashes in smart contracts solves the issue of blockchain storage explosion. And when used in tandem with smart contracts, it enables instantaneous outflows between parties in response to changes in data stored in the blockchain. The paper also offers simulations of the implementation and analysis of the performance. The findings validate that our system improves security for sensitive information, safeguards supply chain data, and meets the needs of real-world applications. Furthermore, it boosts throughput efficiency while reducing latency.

**Keywords**—Blockchain; IPFS; supply chain management; traceability; Ethereum

## I. INTRODUCTION

Everyone is concerned about the nutritional value of the meal they are about to consume. The food supply chain is prevalent, with tedious processes prone to human errors [1]. A blockchain-based supply chain can offer visibility into the movement of commodities by making all transaction records accessible to all network nodes. End-to-end traceability is another feature of the blockchain-enabled food supply chain technology that is helpful in the case of a food investigation or recall. Supply chain management combines actions to transform raw materials into completed goods, increasing customer value and achieving a long-term competitive advantage. The supply chain generally encompasses people, things, initiatives, and organizations involved in transforming raw materials into finished goods, mainly and secondary to

fulfilling customer orders. Although the government has developed national traceability requirements for significant products, fake and subpar interests are still common in the marketplace. Thus, several issues with food security have led to a crisis in consumer confidence, which is also a significant obstacle to national efforts to build a trustworthy society [2].

NP agro-food monitoring program can determine food origin and provide detailed information about food production and distribution. The farm-to-table process includes the production, processing, transportation, and sale of agricultural products. Any fraud in any of the connections above can result in potentially fatal threats to food safety. Therefore, various management solutions have arisen with IoT, automation, and different learnings. These systems can trace the complete procedure autonomously, but they must solve the problem of altered data and food poisoning. The reason for this is that in the conventional data storage procedure, there is always the risk of data being tampered with or lost. Researchers have been experimenting with blockchain technology to store data and protect data security by addressing the issues above [3]. It develops a blockchain-based system for storing agricultural product provenance data. Due to their linkage to IoT devices, agricultural items will automatically transmit whatever data they collect to a network.

Once again, the server will dynamically record the information onto the blockchain following data processing. In this instance, provenance data secure storage is implemented using blockchain technology. As a result, it successfully ensures the authenticity of provenance data. A lot of real-time monitoring data would be produced when many agro items joined the provenance monitoring platform. Blockchain was initially developed for transactions involving digital currencies; the volume of data generated is far less than that of real-time monitoring data. This makes it extremely difficult for the rate of block construction to stay consistent with storing traceability information. Thus, more than a direct application of blockchain technology is needed. We merge the IPFS and blockchain to offer a data storage and query method for a traceability solution for farm products. IPFS is a peer-to-peer decentralized file system with the ambitious goal of connecting all devices that use computers to a single, universal file system. It first suggests a data storage model based on blockchain and IPFS [4]. The submitted video, picture, and sensor data are automatically encapsulated and parsed by this model. The above information is then sent to IPFS, and the blockchain is updated with the associated hash addresses. The blockchain transaction's hash values are then saved in the

database. By requesting the authenticity database hash address of an item in IPFS, customers can access the provenance data using the transaction content from the blockchain.

The article is divided into the following sections: Section II provides a quick review of blockchain technology and terminology, stages, methods, and benefits of traceability. Section III also discusses previous initiatives to incorporate blockchain into provenance systems. Section IV describes a modeling approach that has been presented. Section V goes through the outcomes and performance. Section VI brings the paper to a conclusion.

## II. BLOCKCHAIN FOR TRACEABILITY

Blockchain technology can offer improved tracking of agri-foods from the ranch to your dinner plate. Blockchain technology might greatly minimize the food safety concerns arising from deliberate deception, terrible organization, and being short of guidelines. [5] Blockchain technology will revolutionize how traceability is viewed by overcoming the challenges in Table I.

TABLE I. CHALLENGES IN TRACEABILITY SYSTEM

Challenges
Security and privacy
Credentials and Governing Compliance
Lack of end-to-end visibility
Interoperability
Expiration and Counterfeit Products
Stakeholders Trust management
Conflict of interests
Temperature-controlled Logistics

### A. Functioning of Blockchain

The digital register is a "link" of unique "blocks" information. As new data is added to the network, a new block is created and linked to the chain. To maintain consistency, all nodes (computers) must update their versions of the blockchain ledger. One of the elements that make blockchain so highly secure is the process of adding new blocks [6]. This is due to the requirement that a participating node verifies and validate the validity of new data well before a new node can be recorded to the blockchain. About cryptocurrencies, this can mean proving that brand-new transactions within a block are legitimate and that no coins have been spent more than once.

On the other hand, a single database or worksheet permits revisions to a single version without restriction. Once consensus is reached, the block is added to the ledger, and the underlying records are kept in the distributed ledger [7]. Blocks are safely joined together to form a secure digital chain that extends from the start of the log to the latest addition. Nodes are frequently given fresh amounts of the native currency of the blockchain in appreciation for their work confirming revisions to the linked data.

### B. Merkle Tree

The Merkle binary hash tree of 1987[8] is an essential component of the Blockchain structure. Each node of the

Merkle tree is called a leaf, a separate data item or record. A cryptographic hash function  $H()$ , is used for evaluating a leaf of  $X$  records which in turn has a  $2X - 1$  inner node. A tree with  $X = 2y$  nodes has a complexity of  $y = \log X$ . Also, a single unique root at height  $y = 0$  is the source of the entire hierarchical tree. The claim that the root value  $\hat{R}$  encapsulates the correctness of all the entries in the tree constitutes a statement made to the entirety of the tree. Conforming to each child  $n$  at height  $y'$ ,  $Z$  forms a collection of  $y'$  systems on the whole hierarchy.

$$\hat{R} = HF_t(n, Z) \quad (1)$$

where,  $HF_t$  is a succession of  $y'$  hash function  $H(.)$  calculations. The entire Merkle hash tree is stored by an untrustworthy party to prove, whereas a verifier only stores the root  $\hat{R}$ . The party attempting to verify the authenticity of node  $n$  provides  $Z$  for each node in the tree. Given,

$$\hat{R} = Hf_t(m, Z) \quad (2)$$

Eq. (2) concludes that statement  $n \in \hat{R}$ , where  $n$  is the existing root. Knowing the present  $\hat{R}$ , a steady improvement state occurs from  $\hat{R}$  towards  $R = Hf_t(m', Z_m)$  every time the current node  $m$  reorganized to  $m'$ .

### Hashing

For a continuously increasing value,  $m = \alpha_0, \alpha_1, \alpha_n$ , a hash function  $h(.)$  is applied. This results from a declaration  $\bar{G}$  called Gatherer Hash (GH) to the progression. The process to get  $n$ th hash gatherer  $\bar{G}_n$  is given in eq. (3),

$$\begin{aligned} \bar{G}_0 &= h(\alpha_0), \beta_1 = h(\beta || \alpha_1) \dots \beta_n, \\ &= h(\beta_n - 1 || \alpha_n - 1) \end{aligned} \quad (3)$$

Also, a cryptographic hash function's  $h()$  attribute assures that determining any sequence of numbers other than  $\alpha_0, \alpha_1 \dots \alpha_n$  that yields the current  $GH\beta_n$  is unfeasible.

### C. IPFS

Despite multiple attempts to develop distributed file structures, IPFS has been the initial universal storage structure to achieve high throughput and autonomous delivery on a global scale and on the system level. Peer-to-peer (P2P) networks and blockchain technology allow secure information sharing without a central organization. InterPlanetary File System may replace HTTP (Hyper Text Transfer Protocol). IPFS is a persistent, open-source data storage system. It's used by many to exchange content efficiently. IPFS distributes files throughout the network and uses their cryptographic hashes to identify each item based on its content. IPFS stores file off-chain while referencing blockchain hashes [9]. While IPFS incorporates useful concepts from earlier peer-to-peer systems, its primary contribution is in streamlining, advancing, and integrating tried-and-true methodologies into a single coherent system that is more effective than the sum of its parts [10]. The IPFS constructs an algorithmically authorized data model on top of these, a Merkle Directed Acyclic Graph (DAG) of permanent objects, to aid in effectively disseminating files and file version control. Git13 is a popular system for managing file revisions and distribution. Content addressability, tamper resistance, and

deduplication are all provided by the core IPFS principles [11]. Additionally, the IPFS implements the InterPlanetary Linked Data (IPLD) standards to produce more adaptable, flexible, decentralized data structures that are globally addressable and linkable for many types of data. The Interplanetary Way back is a persistent Web archive that distributes data files into the IPFS network [12]. The CDXJ index is built, and then IPFS is used to disperse the response records' headers and payloads independently. The method has the potential to speed up indexing generally.

### III. RELATED STUDY

To solve the problem of the blockchain's lack of significant data storage, Yang et al. [13] employed Hyperledger as the provenance link that processes the database. Its disadvantages over IPFS data storage include expensive costs, a slow rate of data movement, inadequate privacy, and so on. Additionally, it needs a feedback feature for the customer, preventing retailers from initially accessing the information on product security and other factors. Liao and Xu [14] developed a blockchain monitoring system utilizing intelligent farming and sensor networks to manage tea quality and safety. Using hazard characteristics; they also developed technologies for culinary threat assessment and safety tracing. Xie et al. [15]'s ETH-based traceability of agricultural items used IoT technology to protect data from unauthorized tampering or damage. However, information is saved utilizing blockchain technology at the file storage layer; as a result, network overheads will increase as data volumes do. Bumblauskas et al. [16] integrated IoT and blockchain to track products in real time. Using a Midwestern company's egg supply chain as an example, blockchain technology was implemented from farm to consumer.

A provenance storage solution using blockchain technology would record food progress information in IPFS and analyze such data with the aid of an auxiliary database, claims a study by Hao et al. [17]. Yu and Huang [18] proposed the broiler chicken traceability method by fusing RFID and blockchain technology. The "inverted tooth" design of the chicken claw ring prevents it from being resold. The approach allows intelligent technologies to scan the ring's QR code and access the relevant details and data. Benet [8] advises using the InterPlanetary File System (IPFS) to link all related devices using a single file system. Using a content-addressed strategy, IPFS is a peer-to-peer dispersed file system that provides high-throughput storage space. There are other attempts as well that aim to establish a global file system. The AFS system is installed by Howard et al. [19] and can enhance the capability of cache validation, server process structure, and other things. Additionally, several sizable media file-sharing platforms like Napster, KaZaA, and BitTorrent are developed to hold enormous amounts of data.

### IV. PROPOSED MODELLING

#### A. Non-Perishable (NP) Agro-food Supply Chain Traceability System

Poor server management and maintenance from distributed cloud storage providers or an extremely concentrated variety of cloud providers are the main causes of the present issues

with cloud storage [20]. A backup file's hard disk position might match the original file's hard disk location if the cloud hard drives are centralized, even if both files are kept on the same cloud hard drive. As a result, in the event of a power failure or other issue, the servers malfunction and are unavailable from the outside; the only option is to wait for the servers to recover. Though it is unrestricted, IPFS is a more current Internet technology than the HTTP protocol. It is based on the idea that a file can be divided into numerous bits dispersed throughout the network and obtained from multiple servers consecutively using a P2P network (PPN). If some servers are unavailable, users from outside the network can still connect to the system and access data. Additionally, the network has many backups, even if some nodes' data is completely lost as a result of an error. Data loss, outdated infrastructure, and a lack of user feedback are just a few issues with the conventional centralized public cloud that can be addressed with IPFS's benefits.

High standards for transaction data backup are necessary to ensure the traceability of the supply chains for agricultural products. Since it divides a file into several pieces and disperses them across the network, the IPFS storage system has a more reliable backup capability than cloud storage. The openness of the information kept in IPFS may be ensured by blockchain technology, which is excellent for the supply chain traceability of agricultural products. In this section, we have used the blockchain to track and execute the interactions in the NP agro supply chain products to decrease trust in the central database. We can do this with the help of smart contracts and IPFS transaction records. We can do this by leveraging smart contracts and documenting transactions in IPFS.

#### B. System Model Overview

Numerous parties, including producers, manufacturers, distributors, retailers, logistics, and customers, are involved in the NP agri-food supply chain model. The regulatory framework that governs the traceability system is responsible for distributing both the public and the private keys to each participant.

It is possible to strengthen customers' confidence in the safety of NP agricultural products by providing consumers with complete information on agricultural products via a traceability system. This will allow for increased sales of NP agricultural products. In this article, the relationships between production, manufacture, distribution, retailing, logistics, and sales are disentangled to investigate agricultural goods' transparency. The NP agri-food is planted, transplanted, watered, fertilized, and harvested as part of the production link. Additionally, it requires recording essential data, including details on seedlings, planting techniques, environmental factors, and product transactions. NP commodities are categorized, weighed, packed, priced, and subjected to various operations in the manufacturing unit. It also includes keeping track of data on NP products, production processes, processing parameters, product transactions, and other essential facts. The distribution unit transports the full goods from one place to several others. When handling the delivery of items to the consumer, the shop takes great care. Transportation that is a part of IoT operations is used during production, manufacturing, distribution, and retail operations.



Law enforcement agencies can track down incidents involving how good and safe agricultural products are and identify the principal parties at fault. To certify the integrity and openness of provenance data in farm produce monitoring systems, traceability uses blockchain technology's decentralization, non-tampering, and tracking properties. The production, processing, transportation, and sales of agricultural products may all be monitored with the help of the blockchain-based NP agro products tracking system, which maintains growth data, processing data, logistics details, and sales data. Fig. 1 depicts the structural layout of the blockchain-based monitoring system for NP agro products. The regulator must approve the blockchain traceability scheme, including all parties. The approval follows registration. After registration is complete, the parties involved can transfer individual and NP product traceability data. If you want to check if the data on your chain of custody has been tampered with, you can use the provided a comparison tool to check for validity. Data analysis, a consensus mechanism, key management, smart contracts based on a user's reputation, and key authorization is all part of the system. The most fundamental applications of smart contracts are adding data to blockchains and the subsequent querying of that data. When a request is made, the smart contract immediately begins carrying it out. Businesses may input data in various links, customers can query traceability data, and regulators can keep tabs on it all through the system's platform, which is designed to accommodate a wide range of users. We store the data to IPFS in order to prevent blockchain data explosion. The IPFS hash value is kept in the blockchain once consensus has been reached.

### C. IPFS Encrypted Storage

The blockchain traceability system's current storage option entails immediately writing the traceability data of each system of agricultural goods onto the blockchain. The burden on blockchain's storage grows as there are more nodes, resulting in more transaction data being collected [21]. Due to

the blockchain's unique chain-type topology, users of a single blockchain community can view all the information on the blockchain even when query efficiency is extremely low. To address these issues, this article developed a solution integrating encrypted private data storage and hash storage for public data to enhance the storage style of a blockchain monitoring system for NP agro products. Traceability data for products is just one part of the supply chain's comprehensive dataset, including sensitive information authorized parties may only access. Data confidentiality is a significant concern for rival businesses. Public information may be about the producer, manufacturers, retailers, logistics and their reputation, product details, date of manufacture, price, and provenance. In this work, Algorithm 1 depicts the data entry to the blockchain in protecting traceable information, in which sensitive data is encrypted using a smart contract and then stored on the blockchain alongside the corresponding hash value of public data.

---

#### Algorithm 1: Double Encryption

---

```

Initialize Stakeholders ID, Public_data, Sensitive_info, Keys.
Stakeholders ID → SID
If
    Sensitive_info! = Null then do
        Generate a Random key (KA)
        Sensitive_info → Encry (GCM (Sensitive_info, KA))
        Encry_key → Encry (Key_Encryption_Method
        (KA, PuK))
        (Sensitive_info + Public_data) → IPFS
        Hash (Sensitive_info) + Hash (Public_data) → BC
    End If
If
    Sensitive_info == Null then
        Hash (Public_data) → BC
    End If
End
Result: The hash value stored in the blockchain.
    
```

---

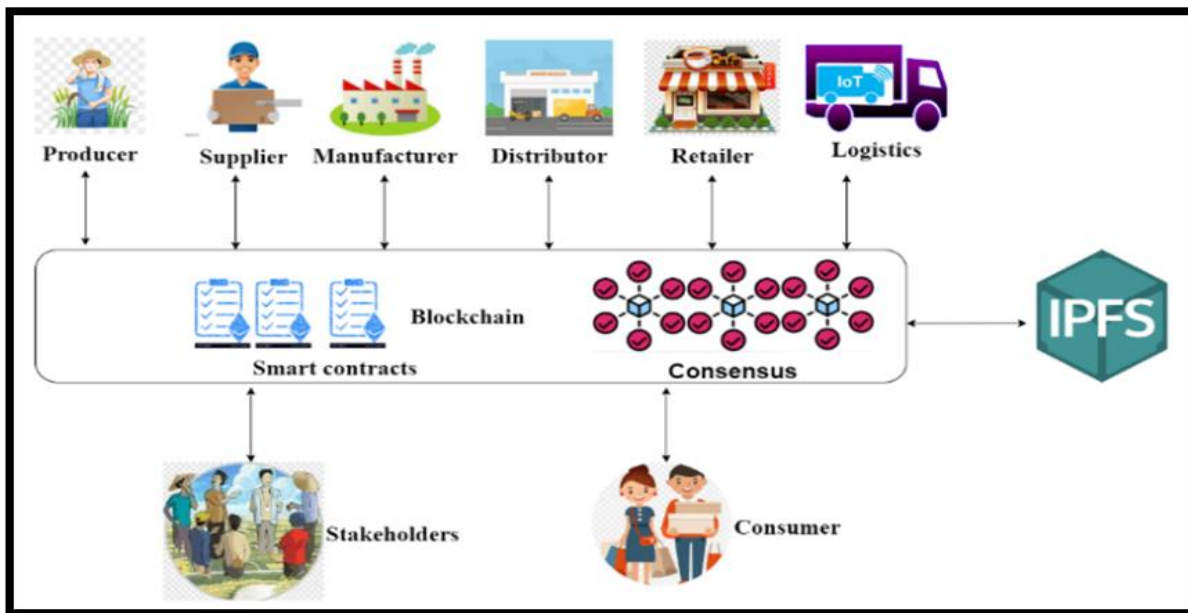


Fig. 1. Blockchain-oriented traceability mode.

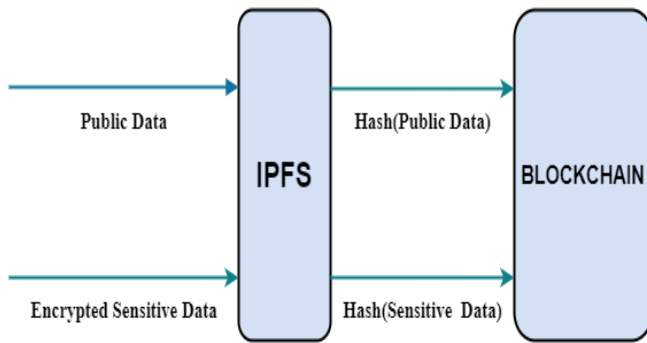


Fig. 2. Storage of public and sensitive data.

The AES encryption technique uses the Galois/Counter Mode (GCM) to encode sensitive data, 'Sensitive\_Info'. The necessary Key, 'K<sub>A</sub>' is chosen at random by a random generator using the smart contract, which also creates as well as uploads encoded substitution cipher to the network. The Key is encrypted using the Key Encryption Method (ECC) to guarantee its security. The viewing node was approved by the encrypted Public Key, 'Pu<sub>K</sub>'. Sensitive data as well as open data are transferred to IPFS. A pair of keys made up of the Encrypted Key and the Public Key of the Trusted Observing Nodes are kept in the current picture of the smart contract and sent to the ledger as IPFS hashes as shown in Fig. 2. The existing node's private key, 'Pi<sub>K</sub>' is used to decipher the blockchain's encoded key in order to acquire the original Key, which is then used to decipher the sensitive data and sight it when the appropriate nodes access the private data.

## V. RESULTS AND DISCUSSIONS

Ethereum is a blockchain open-source platform that is used for simulations. We create and test smart contracts using the Remix Integrated Development Environment (IDE), Ganache, Truffle, and Metamask. The smart contracts are used and tested on the Ethereum Ropsten test networks. The script was created using the Remix IDE and Solidity version 0.8.7. The system requirements include an x64-based processor, Win10 Pro, a 64-bit operating system, an Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz, and IPFS version 0.8.0. Ganache does nevertheless distribute a certain amount of cryptocurrency to virtual accounts. After each transaction, the cryptocurrency is removed from the contract's designated account. In Ganache, every account has a unique address and private key. The truffle is the framework. Through the Metamask, a bridged browser plugin, Ganache, and the Remix IDE may communicate with one another.

Through IPFS, the file is uploaded to the blockchain. It displays a ledger with a number of records. The key that represents each transaction is used to identify each record. The record is a file's many hashes. Each entry starts with a key, then a record that includes the multi-hash value of the file, the ID of the network node reading the file, and the time stamp. Due to the network's addition of IPFS hash addresses, access to files on blockchains using IPFS is constrained. This restricts access to the IPFS system and its associated hash addresses to legitimate users only. This lowers the number of handlers who have admittance to the files. Also, lessen illegal admittance to

the file as well. Since user activity is not recorded on the IPFS network, users are free to view any files without fear of repercussions. This experiment places a barrier on unwanted access due to the immutable logging of user file access.

### A. Consensus

The consensus mechanism assures blockchain nodes have the latest blocks. Popular ones include Proof of work, Proof of stake, Delegated Proof of Stake, and Practical Byzantine Fault Tolerance [22]. By facilitating competition between dispersed nodes in terms of computational power, POW increases data consistency and integrity, which is helpful in fending off Sybil attacks. The researchers created a POS approach that doesn't require processing resources and a DPOS mechanism to reduce power consumption. PBFT can reach a consensus with a few malicious nodes. Signatures, verification, and hashing prevent tampering, counterfeiting, and message repudiation and made BFT polynomial. PBFT's high consensus cost renders it unsuitable for public chains. Fewer-node consortiums or private chains operate better. The consortium chain is the core network, while PBFT is the consensus method.

### B. Performance and Evaluation

Here, we measure how long it takes to verify a transaction by evaluating all the transaction data from the moment the validator first receives it until the transaction is authenticated. On the Ethereum Ropsten testbed, we assess the latency and throughput of our implementation. The approximate transaction validation time is given for all transaction types, including stakeholders.

Key extraction in ECC encryption is as simple as producing a random integer between two specified values. Any positive integer within the range is valid as an ECC secret key. It is possible to minimize the time spent waiting by taking two different approaches when computing keys throughout decryption and encryption. ECC provides less delay than the alternative by processing keys in two stages. Since the ECC key is shorter, it consumes less computing capacity while still providing secure protocols for authorized mutual authentication.

Some of these are reasonable despite the large additional operations required, such as validating signatures and hash data to and from IPFS, confirming permissions in the control list, and performing several other overheads. Transactions take the longest to complete since the qualifications of both the buyer and the vendor must be confirmed. The effectiveness of the blockchain network is examined in this section regarding various transaction rates. As a result, testing involved using transaction counts of 50, 100, 150, 200, 250, and 300 per second. By changing the number of transactions occurring, the impact of having numerous transactions active in a blockchain was examined.

Standard operations and query transactions were both put to the test in every situation. The number of total transactions has been adjusted to facilitate the development of a more comprehensive understanding of how the amount of all transactions influences the throughput and latency of the blockchain. Each transaction on the blockchain has its

throughput and latency recorded and logged for future reference. This information is stored in a distributed ledger. Fig. 3 and 4 presents the calculations' results on the specifics of the delay and throughput measurements.

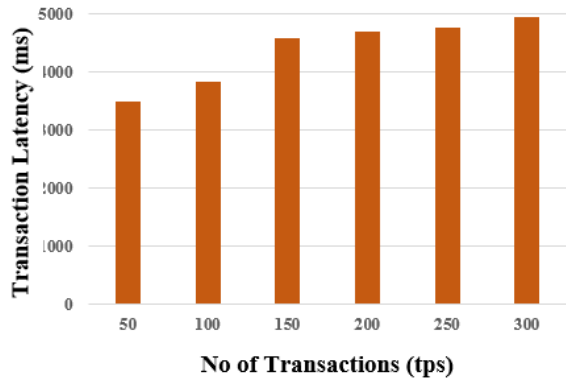


Fig. 3. Transaction throughput of the traceability system.

TABLE II. AVERAGE LATENCY AND THROUGHPUT OF THE TRACEABILITY SYSTEM

	Ethereum (Ropsten)
Average Latency	3927.32 MS
Average Throughput	52.21 TPS



Fig. 4. Transaction latency of the traceability system.

Average delay and average throughput are determined from test results on the Ropsten test networks and are shown in Table II. Calculations are done for various transactions, including registration, the record of trace data, approvals, querying of recalls, etc.

### C. IPFS Read Write Latency

While reading files over the network, IPFS reads them more quickly than it writes or adds them. A 0.256 MB file is typically received in around three seconds and written to a remote IPFS gateway in about five seconds. Because it caches the data locally after the initial transmission, IPFS is substantially faster when requesting the same file a second time. We also tested the performance of IPFS on its own to compare the outcomes of the combined systems. Files were

sent between local workstations, and the server latency fluctuated. The impact of delays is more significant on smaller files than on bigger ones. The following is a performance analysis of scaling-out IPFS-attached storage. Table III displays their reading and writing results for different file sizes and latency.

TABLE III. IPFS READ WRITE LATENCY

Read Latency	5MS	10MS	20MS	25MS
25MB	0.395	0.511	0.755	0.841
100MB	1.453	1.505	1.93	1.82
1GB	13.26	13.24	14.14	15.11
Write Latency	5MS	10MS	20MS	25MS
25MB	0.443	0.474	0.741	0.891
100MB	1.597	1.694	1.893	1.99
1GB	14.10	14.73	15.32	15.89

### D. Comparisons

Our system has the following benefits: a better standard of decentralization; better network resilience; more integrity in data communication; and high reliability in data processing, according to the assessment of findings. The time-suggested technique enables more transactions to be added to the chain at the same rate, serving more users and enhancing overall scalability. We are evaluating our solution in comparison to several traceability systems.

TABLE IV. COMPARATIVE RESULTS

Features	Conventional Traceability	Blockchain Traceability
Traceability	Yes	High
Supervision	Medium	High
Credibility	Less	High
Storage Mode	Medium	Medium
Data sharing Scope	Medium	High
Scalability	Less	Medium
Query Rate	High	Medium

Table IV contains comparable outcomes. The performance of the decentralized and blockchain-based traceability systems is lower than that of the proposed method. All designs can track data. The proposed system cannot be altered in any way, making it superior to centralized systems and other approaches. Our method is decentralized, scalable, and protects users' privacy at the same time. This method reduces the amount of data generated compared to conventional blockchain technology.

## VI. CONCLUSIONS

In this study, we developed and assessed a system for tracking agricultural products using blockchain's tracing and anti-tampering features. We also examined the system's architecture for querying and storage. The problems of the blockchain traceability system's enormous data overload

constraint and the lack of private security is addressed, and a suggestion for on-chain and off-chain file storage is made. The public data consumers see in the supply chain is stored in the IPFS, whose hash value is uploaded to the ledger system. For discussing with related firms, the blockchain maintains private data that has been encrypted using an encryption method. The storage approach described in this model integrates the existing scenario, considering the necessity for public scrutiny of supply chain public data and the requirement that private corporate information is encrypted, and minimizes the data burden on the chain. After the consumer scans the Tracking number for public database information, the system validates the information using the relevant block to verify if the product information has been updated. Blockchain technology is expected to develop in the direction of numerous test beds, platforms, and sharding. We will continue looking into the technology tying together diverse platforms and a cutting-edge consensus mechanism suitable for traceability in future studies.

#### ACKNOWLEDGMENT

The authors would like to express their gratitude to everyone who helped them write this article and provided data for this research.

#### REFERENCES

- [1] K. Sornalakshmi, S. Sindh, G. Sujatha, and D. Hemavathi, "An architectural framework of a Decision Support System (DSS) to increase the returns of small-scale farmers in Kanchipuram District, India," *EAI Endorsed Trans. Energy Web*, Vol. 0, No. 0, 2018, pp. 163978, <https://doi.org/10.4108/eai.13-7-2018.163978>.
- [2] L. Zhang, W. Zeng, Z. Jin, Y. Su and H. Chen, "A Research on Traceability Technology of Agricultural Products Supply Chain Based on Blockchain and IPFS," *Security and Communication Networks* 2021, 2021, <https://doi.org/10.1155/2021/3298514>.
- [3] F. Tian, "A supply chain traceability system for food safety based on HACCP, blockchain & Internet of things," in 2017 *International Conference on Service Systems and Service Management*, 2017, <https://doi.org/10.1109/ICSSSM.2017.7996119>.
- [4] M. K. Pawar, P. Patil, and P. S. Hiremath, "A Study on Blockchain Scalability," in *Advances in Intelligent Systems and Computing, Singapore: Springer Singapore*, 2021, pp. 307–316, [https://doi.org/10.1007/978-981-15-8289-9\\_29](https://doi.org/10.1007/978-981-15-8289-9_29).
- [5] S. Balachander and A. Murugan, "Challenges and Opportunities in the Advancement of Privacy and Security of Blockchain Technology," *Telematique*, Vol. 21, No. 1, pp. 6651-6659, 2022.
- [6] Shanthi and K. Venkatesh, "An analysis of various techniques in blockchain applications," in 2022 *6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2022, <https://doi.org/10.1109/ICICCS53718.2022.9788137>.
- [7] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system." *Decentralized Business Review* (2008), pp.21260,2008, (White paper).
- [8] R.C.Merkle, "A digital signature based on a conventional encryption function." In *Conference on the theory and application of cryptographic techniques*, Springer, Berlin, Heidelberg, 1987, pp. 369-378, [https://doi.org/10.1007/3-540-48184-2\\_32](https://doi.org/10.1007/3-540-48184-2_32).
- [9] E. Politou, F. Casino, E. Alepis and C. Patsakis, "Blockchain mutability: Challenges and proposed solutions." *IEEE Transactions on Emerging Topics in Computing*, Vol. 9, No. 4, 2019, pp.1972-1986, <https://doi.org/10.1109/TETC.2019.2949510>.
- [10] J. Benet, "Ipfes-content addressed, versioned, p2p file system." arXiv preprint arXiv:1407.3561, 2014, <https://doi.org/10.48550/arXiv.1407.3561>.
- [11] C. Patsakis and F. Casino, "Hydras and IPFS: a decentralised playground for malware." *International Journal of Information Security*, Vol. 18, No. 6, 2019, pp.787-799, <https://doi.org/10.1007/s10207-019-00443-0>.
- [12] S. Alam, M. Kelly, and M. L. Nelson, "Interplanetary Wayback: The permanent web archive," in *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries*, 2016, <https://doi.org/10.1145/2910896.2925467>.
- [13] X. Yang, M. Li, H. Yu, M. Wang, D. Xu, and C. Sun, "A trusted blockchain-based traceability system for fruit and vegetable agricultural products," *IEEE Access*, Vol. 9, 2021, pp. 36282–36293, <https://doi.org/10.1109/ACCESS.2021.3062845>.
- [14] Y. Liao and K. Xu, "Traceability system of agricultural product based on block-chain and application in tea quality safety management," *Journal of Physics: Conference Series*, Vol. 1288, No. 1, 2019, pp. 012062, <https://doi.org/10.1088/1742-6596/1288/1/012062>.
- [15] C. Xie, Y. Sun, and H. Luo, "Secured data storage scheme based on block chain for agricultural products tracking," in 2017 3rd International Conference on Big Data Computing and Communications (BIGCOM), 2017, <https://doi.org/10.1109/BIGCOM.2017.43>.
- [16] D. Bumblauskas, A. Mann, B. Dugan and J. Rittmer, "A blockchain use case in food distribution: Do you know where your food has been?" *International Journal of Information Management*, Vol.52, 2020, pp.102008, <https://doi.org/10.1016/j.ijinfomgt.2019.09.004>.
- [17] J. Hao, Y. Sun, and H. Luo, "A safe and efficient storage scheme based on blockchain and IPFS for agricultural products tracking," *Journal of Computers*, Vol. 29, 2018, pp. 158–167, doi:10.3966/199115992018122906015.
- [18] W. Yu and S. Huang, "Traceability of food safety based on block chain and RFID technology," in 2018 *11th International Symposium on Computational Intelligence and Design (ISCID)*, 2018, DOI 10.1109/ISCID.2018.00083.
- [19] J. H. Howard et al., "Scale and performance in a distributed file system," *ACM Trans. Computer Systems*, Vol. 6, No. 1, 1988, pp. 51–81, <https://doi.org/10.1145/35037.35059>.
- [20] K. Venkatesh, L. N. B. Srinivas, M. B. Mukesh Krishnan, and A. Shanthini, "QoS improvisation of delay sensitive communication using SDN based multipath routing for medical applications," *Future Generation Computer Systems*, Vol. 93, 2019, pp. 256–265, <https://doi.org/10.1016/j.future.2018.10.032>.
- [21] J. F. Galvez, C. Juan, and J. Mejuto, "Future challenges on the use of blockchain for food traceability analysis," *TrAC Trends in Analytical Chemistry*, Vol. 107, 2018, pp. 222–232, <https://doi.org/10.1016/j.trac.2018.08.011>.
- [22] K. Harshini Poojaa and S. Ganesh Kumar, "Scalability Challenges and Solutions in Blockchain Technology," in *Inventive Computation and Information Technologies*, Singapore: Springer Nature Singapore, 2022, pp. 595–606, DOI: 10.1007/978-981-16-6723-7\_44.

# Implementation of Flood Emergency Response System with Face Analytics

E.Mardaid<sup>1</sup>, Z. Zainal Abidin<sup>2\*</sup>, S. A. Asmai<sup>3</sup>, Z. Abal Abas<sup>4</sup>

Master Student<sup>1</sup>

Fakulti Teknologi Maklumat Dan Komunikasi, Universiti Teknikal Malaysia Melaka<sup>1, 2, 3, 4</sup>  
Durian Tunggal, Melaka, 76100, Malaysia<sup>1, 2, 3, 4</sup>

**Abstract**—Disaster management system is developed to monitor flood, tsunami and earthquake, which is effectively preparing for and responding to the disaster. In fact, Malaysia is building the emergency response system for managing the flood disaster since it is highly occurred. However, the current flood emergency response system has limitations, which, has no integration that data entered into spreadsheets and transferred to different storages, and, data analytics that assist in data collection and decision-making. Even though flood emergency response system have been improved, which introducing sirens and loudspeakers to reach flood victims but still difficult to access the basic facilities and receive flood responses. Therefore, this study implements a flood emergency response system with face analytics to assist data acquisition, which analyze flood victim's faces using CCTV infrastructure with HOG algorithm that incorporated with a dashboard. The dashboard categorizes the number of flood occurrence, maximum flood period (days), the number of displaced flood victims and loss assessment. Findings have shown that the dashboard helps the enforcement agencies to implement the real-time information about flood victims. Based on the number flood frequency occurrence in Malaysia from 2017 to 2021, the percentage produced was 27%, 19%, 12%, 19% and 23%. Moreover, the duration of floods has been decreased from 30% to 17% in 5 years that shows the flood emergency response helps the Malaysia government to improve the infrastructure. The significant of this study beneficial to the local enforcement unit and evacuation centers in identifying the flood victims.

**Keywords**—Disaster management system; flood emergency response system; flood emergency response system with face analytics; flood emergency response system with face analytics using HOG algorithm; flood emergency response system dashboard

## I. INTRODUCTION

Malaysia is located in Southeast Asia, bordering Thailand, Brunei and Indonesia, which geographically is located outside the ring of fire and south of major hurricanes. However, Malaysia is often affected by other natural disasters such as floods, landslides, haze, earthquakes, droughts, tsunamis and other man-made disasters. Floods cause the most frequent and severe damage every year, causing loss of life, disease, property, and other damages [1]. In Malaysia, flooding has become a major disaster problem. From July 2012 to January 2019, Malaysia's population was the most highly affected by floods among ASEAN members [2].

As a result, in 2015, the National Disaster Management Authority (NADMA), under the Prime Minister's Office, became the premier disaster management agency for disaster management efforts in the national and international level. The organizational structure of disaster management continues at three levels: federal, state, and county [3].

NADMA provides Flood Early Warning Systems (FEWS) that divided into five classifications as pre-internet, internet of content, internet of services, internet of people and IoT (internet of things) that warn the public about disaster risks going to happen [4], [5]. Due to the unpredictable estimation of flood occurrence, several methods have been proposed to enhance the FEWS.

The first solution is to create a monitoring tool that observes the occurrence of floods and warns victims before trap in the flood area. The importance of this tool is to reduce the number of flood victims in the flood areas, enable rapid evacuation of victims, and prevent loss of life and save more flood victims from flood injuries [6].

The second solution proposes a new dashboard to provide latest information about flood victims, which comes from face images of flood victims that been capture from CCTV. The propose system called as flood emergency response system with face analytics will verify the captured face image from CCTV with the stored face image using HOG algorithm. The face analytics acts as a back end engine for the proposed system.

The rest of the paper is organized as follows: Section II describes the literature review of flood and its causes. Section III explains on methodology used to design the propose system. Section IV explains about the architecture and techniques. Section V explains about the results and discussion. Section VI summaries the implementation of the flood emergency response system with face analytics for better performance in identifying flood victims.

## II. LITERATURE REVIEW OF FLOOD

Flood is a dynamic phenomenon with wide geographical and temporal variation. Furthermore, changes in the flood system, such as levee breaches or the adoption of emergency measures, might alter flow patterns and cause unexpected deviations from pre-planned contingencies. Even though floods

\*Corresponding Author.

may affect human activities and assets in a number of ways, disaster management must be aware of which regions are flooded when and how intensely. In contrast, flood impact assessment is mainly concerned with the greatest flooding extent and maximum inundation depths that occurred during the storm [7].

In Malaysia, flooding is categorized into three types, which are river floods, pluvial floods, and coastal floods.

Fig. 1 depicts a river flood, also known as a fluvial flood, which occurs when the water level in a river, lake, or stream rises and overflows onto the banks, beaches, and neighboring land. Significant rain or snowmelt may have caused the rise in water levels.

Fig. 2 shows a pluvial flood occurs when a heavy rainfall event generates a flood that is not produced by an overflowing water body. A common misconception about floods is that you must be near a body of water to be impacted. Nonetheless, pluvial flooding may occur everywhere, urban or rural, and even in areas where there are no adjacent bodies of water. There are two forms of tributary flooding:

- When an urban drainage system gets overflowing, water rushes onto streets and surrounding structures, resulting in surface water flooding. It happens gradually, giving people time to flee to secure locations, and the water level is usually shallow (rarely more than 1 meter deep). It does not endanger people's lives immediately, but it has the potential to cause significant economic harm.
- Flash floods are categorized as a violent, high-velocity torrent of water generated by heavy rain falling in the region or on surrounding higher terrain in a short period of time. A fast release of water from an upstream levee or dam creates a dangerous and catastrophic situation. This is because of the force of the water produce the hurling debris.

Fig. 3 explains about coastal flooding or named as storm surge, which is the inundation of land near the coast by seawater. Windstorms that coincide with high tide (storm surge) and tsunamis are common causes of coastal flooding. Storm surge occurs when high winds from a windstorm cause water to wash ashore; it is the most prevalent cause of coastal flooding and, in certain cases, the deadliest threat associated with a windstorm. Windstorms during high tide cause damaging storm surge floods; windstorms during low tide cause disastrous storm surge floods. In this type of flood, water overwhelms low-lying regions, resulting in significant loss of life and property. Other factors that influence the strength of a coastal flood include the windstorm's power, magnitude, speed, and direction. Furthermore, the onshore and offshore geographies are another critical factor of coastal flood that bring severe damages. Coastal flood models employ this information, as well as data from previous storms in the region, to determine the likelihood and severity of a storm surge.

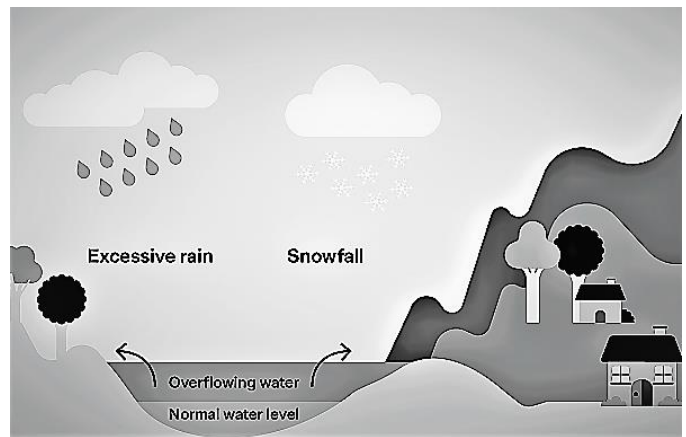


Fig. 1. River floods (fluvial floods).

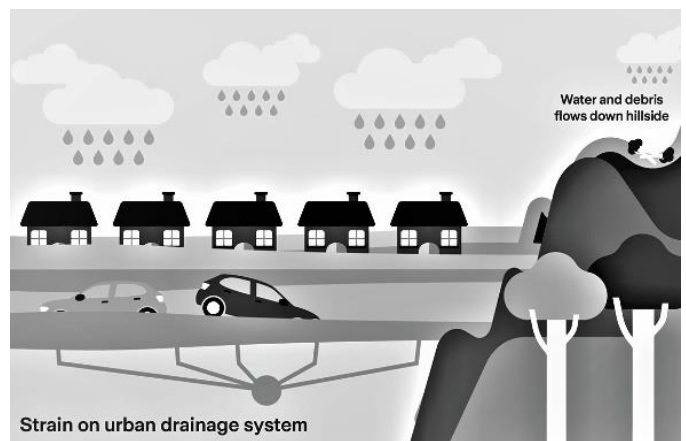


Fig. 2. Pluvial floods (flash floods and surface water).

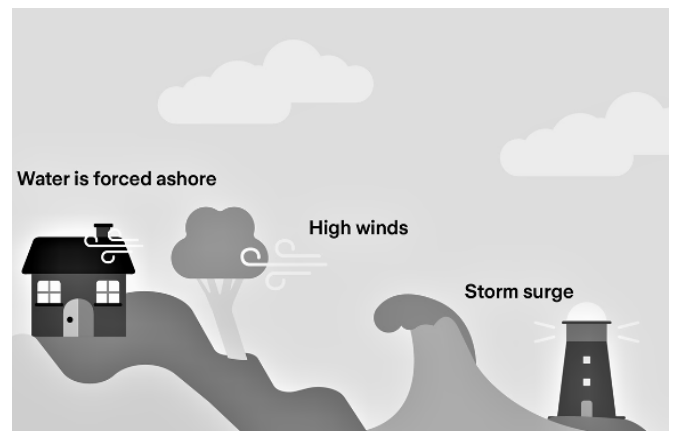


Fig. 3. Coastal flood (storm surge).

#### A. Causes of Flood in Malaysia

Regardless of the type of flood exist; it is always brought on by the same basic event and the effect typically harmful. Meanwhile different types of floods have different causes, most floods are caused by one of the following events [8] [9]:

1) *Heavy rain*: This is the most common source of floods. When there is too much rain or it falls too rapidly, it literally does not have anywhere to go. This can result in floods, including flash flooding.

2) *River flooding*: River flooding is not necessarily caused by heavy rain. As previously said, river flooding can occur when there is debris in the river or dams that restrict the flow of water.

3) *Broken dams*: Older infrastructure may fail when heavy rains fall and water levels increase. When dams collapse, torrential rains fall on unsuspecting populations.

4) *Storm waves and tsunamis*: Hurricanes and other tropical storms may cause sea levels to rise and bury normally dry coastal areas in many feet of water. Tsunamis, on the other hand, are gigantic waves caused by earthquakes or volcanic eruptions beneath the sea. As these waves approach inland, they acquire height and have the potential to drive a huge volume of water inland in coastal areas.

5) *Channel with cliffs*: When there is fast runoff into lakes, rivers, and other basins, flooding is prevalent. This is frequent in rivers and other bodies of water with steep banks.

6) *Lack of vegetation*: Vegetation can help to slow down runoff and prevent flooding. When there is a lack of vegetation, there is little to prevent water from rushing back and flooding rivers, lakes, and streams.

7) *Moon "Wobble"*: The moon wobble is caused by fluctuations in the moon's elliptical orbit and the related gravitational pull on Earth.

### III. FLOOD MANAGEMENT SYSTEM

Flood management system (FMS) has been used across the world to accomplish the goals of the Framework for Disaster Risk Reduction and the Sustainable Development Goals, which the Disaster Risk Reduction (DRR) strongly recommends. The four fundamental aspects of a comprehensive FMS are risk information, monitoring and forecasting, warning, dissemination, and communication, and reaction capabilities. The complexity of operational FMS varies based on the data given, the technology employed, and the know-how [1].

There are substantial differences in maturity between FMS in developed nations that have the financial resources, technological infrastructure, and human resources, and those in poor countries who have less sophisticated FMS. Fortunately, advances in remote sensing, artificial intelligence (AI), information technology, and social media are creating substantial changes in FMS processes, allowing all FMS to obtain enhanced capabilities. These technologies enable underdeveloped nations to overcome the technological hurdles that FMS have previously encountered. The difficulties that FMS confront, as well as technological improvements and their consequences for flood forecasting and disaster response is explored [4].

Flood risk can be assessed by reviewing the historical record of flooding in a certain location, monitoring the people and property damaged by the flood, and use predictive modelling, expertise, or a combination of the above [16]. It is

also necessary to understand the possibility and repercussions of floods. The availability of extended hydrological data series frequently limits the utility of previous flood data. Understanding risks necessitates costly maps and research, which must be updated as dynamic cities expand. Collecting geographical data sets such as topography, land use, soil, and exposure is the first step in predictive modelling. We also require competent individuals to simulate and evaluate the outcomes. As a result, of these challenges, risk awareness is lacking in the majority of FMS covered locations.

To effectively estimate the development of flood water levels, FMS operators must have a thorough understanding of the present and historical values of important hydraulic parameters in the watershed upstream of the area of interest, as well as their likely future evolution. Water levels, discharges, snowfall, precipitation, and temperatures are all often observed factors. The present and historical values of these parameters are available via a monitoring system, which is typically comprised of a network of ground stations supplemented by occasional field surveys. Ground stations typically measure variables of interest at a single location, providing FMS operators with only limited information on the state of the watershed, especially if the monitoring network has insufficient geographic coverage. Approximately 75% of the flood forecasters polled stated that their river basins lack sufficient gauging stations for rainfall, water level, and streamflow information. Fifty percent of the FCCs who answered disclosed that their measurement equipment, gauges, and data transmission devices are outdated. Another challenge is data transfer from stations to forecasting centers, as a considerable proportion of stations in the developing world rely on human observers, hindering the accuracy and timely transmission of data. As a result, most poor nations struggle to capture the amount, distribution, and change of essential variables like precipitation and streamflow during extremes. Weather radars and remotely sensed rainfall have the potential to improve watershed monitoring by providing geographical, real-time, or near real-time data. However, the data they supply is of lesser quality than observed data, and these approaches are too young to have long enough time series to develop credible hydrological models [17].

The acquired hydrological data should ideally be kept in a database and analyzed in real-time using hydraulic models. Not all nations have a centralized, continually updated database. Data is sometimes stored in spreadsheets for quality assurance before being transferred to a central database a few days or months later. One of the most significant issues confronting operational systems is a shortage of technical competence and human resources. CTFs should have people with considerable flood forecasting training and a large enough forecasting team to successfully deliver timely warnings 74% of flood forecasters, however, admit that their centers lack expertise and personnel capable of integrating data, formulating predictions, and distributing information. This might be owing to a lack of expert specialists in job areas, as well as an increase in the demand for disaster recovery and rescue operations after large floods. The shortage of committed permanent personnel in underdeveloped nations is a key impediment to the efficient operation of FMS. Overall, study results suggest that

forecasters have mostly technical expertise but lack understanding of flood vulnerability assessment, warning communication, and downstream resilience, including evacuation preparations [18].

The rapid change of meteorological variables, particularly precipitation, is a critical input for flood forecasting. Meteorologists employ complex computer models to forecast the weather, in addition to ground and satellite data of the land and atmosphere. Numerical weather predictions (NWP) use current weather measurements and combine them using computer models to estimate the weather's future condition. Understanding the present status of the land, ocean, and atmosphere, which effects modelling initialization at local, regional, and global scales, is a significant problem for quantitative precipitation forecasting. In addition, forecasters in industrialized nations have access to fine resolution modelling using ensemble projections to reflect variability and uncertainty in forecasts. Nationally operated meteorological stations in industrialized nations give worldwide predictions from global NWP systems available in underdeveloped countries, such as the Global Forecasting System. This output, however, has a relatively limited spatial resolution. National authorities should scale these forecasts down to greater geographic resolutions [5].

Some nations, however, are unable to make high-resolution predictions due to a lack of critical skills and resources. Even if the prediction is lowered, residual inaccuracies and a fundamental lack of weather forecasting capabilities may result in the failure to detect approaching severe floods or false positives. Even the weather prediction is intelligent. The quality of the data assimilation technique used to derive the beginning conditions of the hydraulic model influences those forecasts. It is also influenced by the hydraulic model configurations described in the model structure, physical process simplification, and calibration quality, as illustrated in Fig. 4, which shows the impact on the performance of the flood forecasting model. Almost half of study that claim the model utilized to produce early warnings is insufficiently accurate or sophisticated for this purpose. As a result, anticipated hit rates varied among model systems and river basins in various climatic zones [19].

Therefore, the new flood management system contains the cutting-edge technology that can easily discover victims as shown in Fig. 4. Moreover, the method of biometric authenticates the specific victim in order to confirm that victim identity. As a logical consequence, physical papers are increasingly becoming obsolete and being replaced with biometric identification. It may also be claimed that current technology has advanced significantly such as allowing consumers to unlock mobile devices using fingerprints or transfer money using voice commands. In order to develop a victim's identification, a biometric analyze unique biological and physiological traits is used. Fingerprints, face, voice, iris, and palm or finger vein patterns are the most frequent biometric identifiers. Banks, for example, require the biometric data in order to perform their numerous services remotely. People have to go to a branch to create an account or make a loan, but now user access numerous services using the newest technology such as a phone, i-watch, laptop, and so forth.

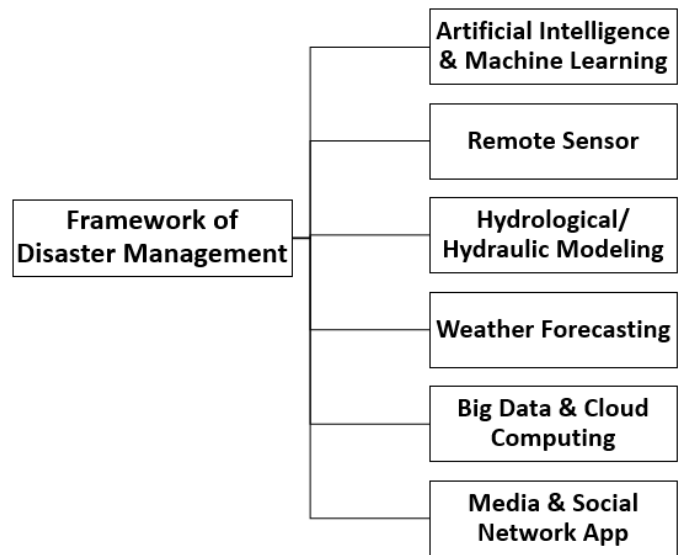


Fig. 4. Potential of new technologies for flood management system [15].

Fig. 5 states about the physical biometrics, in contrast to behavioral biometrics, which relies on measures of the human body for identification or verification. Face geometry, fingerprints, certain areas of the eye, vein patterns, and other physical characteristics are examples of this. It is simply said that physical biometrics substitutes "things you know" (passwords and PINs) with "things you are". This sort of biometrics employs a hybrid approach of analysis that combines human experience with the power of artificial intelligence to provide findings with a high degree of accuracy. The most prevalent approach for authentic user identification nowadays is static physical biometrics. They are utilized by the vast majority of firms (almost 80% globally) that collect and retain physical biometric data to validate identities for a variety of purposes [10]. Thus, the physical biometric is suitable for flood victim identification.

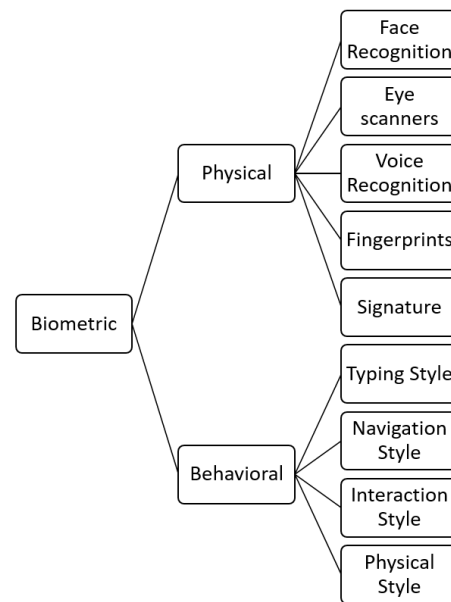


Fig. 5. Types of biometric template.



#### IV. FLOOD EMERGENCY RESPONSE SYSTEM WITH FACE ANALYTICS DASHBOARD (FERSFAD)

The design used for FERSFAD is similar to the laboratory attendance dashboard website based on face recognition system [11] and the technique used for the FERFAD is HOG. CCTV is an idea to overcome problem in contactless criteria that help for a long distance detection [11] [12].

The design of Flood Emergency Response System (FERS), is divided into two parts: facial analytics which is one of the physical biometric and the dashboard that using Power BI. In the facial analytics contains the face analytics components, which are the back end as well as front end of the system of the flood emergency response system dashboard (FERSFAD). The FERSFAD system uses a CCTV camera to send video footage to the computer server that acts as a database for storing image records and calculating face templates using HOG algorithms.

Facial biometrics remains the favored biometric standard due to its ease of deployment and implementation for long distance coverage. In fact, there is no direct contact with the end user and reduces the number of touchpoints on the surface. Furthermore, the face detection and face matching processes is vital for rapid verification or identification, which aids in the search for missing individuals.

In the FERSFAD system, the dashboard creation provides standard requirements of development. Moreover, a huge amount of data from other centers is utilized to give accurate information about the flood. In addition, a quick reference of information about the flood easy to be share among the society and prevent from fake flood information. The propose dashboard has the potential to improve society by boosting safety and security, preventing crimes, and decreasing victim contact.

Therefore, the process starts with capturing the image of the flood victim at the evacuation center. The image of the flood victim is captured using CCTV that has been located at the area of evacuation places [12]. CCTV assists in the data collection phase either the data collection is done during day or night. In the server, the HOG algorithm calculates frames of facial images of the flood victims.

In the enrolment phase, the HOG algorithm involves several steps, which start with Step 1 that is image preparation and color normalization. In the Step 1, the original image is resized to be [8 x 16] metrics and [128 x 64] metrics and the color normalization process transform the resized metrics to be greyscale format. In Step 2, the resized and normalized image metrics are calculated using magnitude and angle of gradients in order to make the image appear closer, which called as face structure. Furthermore, Step 3 implements binning of face structure with gradient calculation using spatial formula that is vital for the matching process. The binning process transforms block normalization into a single-cell normalization that is done in the Step 4. Step 5 involves the HOG feature vector and Scale Invariant and Feature Transform (SIFT) algorithm to search the similar features and save into a database [20].

The HOG algorithm has several methods of detection and matching phase for flood victim faces. The matching technique

involves SIFT process that find the similar features in victims' face captured.

Formula for HOG is magnitude ( $\mu$ ) =  $\sqrt{[(Gx)_2 + (Gy)_2]}$  and angle( $^\circ$ ) =  $\tan^{-1} (Gy / Gx)$ . Benefits of using HOG algorithm in the FERSFAD system is using the SIFT to extract local features, whereas HOG is used to extract global features [13].

To implement the FERSFAD system, the flowchart is designed in estimating overview of the whole process. Fig. 6 illustrates about the flowchart. In the flowchart, it begins with image stored in system database. The image stored in the database is collected from the CCTV that has been collected earlier. The data comes from various units such as evacuation center (indoor) and (outdoor). The CCTV camera detects the victim's face and compare with victim's face image in the database. Once the victim's face match, it proceeds to the recognition process. If the victim's face not matches with the image in the database, person-in-charge of the system needs to add victim's details in system database to update the database. The dashboard displays the detail information about the flood victims.

The programming languages used for the FERSFAD system is Pycharm, which is a Python-specific Integrated Development Environment (IDE) that provides a wide range of necessary tools for Python developers [12]. These tools are tightly integrated to offer a pleasant environment for productive Python, web, and data science development at the back-end of the system. The dashboard connects with phpMyAdmin because it is a web-based utility for managing the MySQL database [12],[13]. Moreover, phpMyAdmin provides an easy-to-use graphical interface for running SQL commands and performing SQL operations. Therefore, at the front-end of the system, the Power BI platform is used to visualize the dashboard for representing the information of the flood victims.

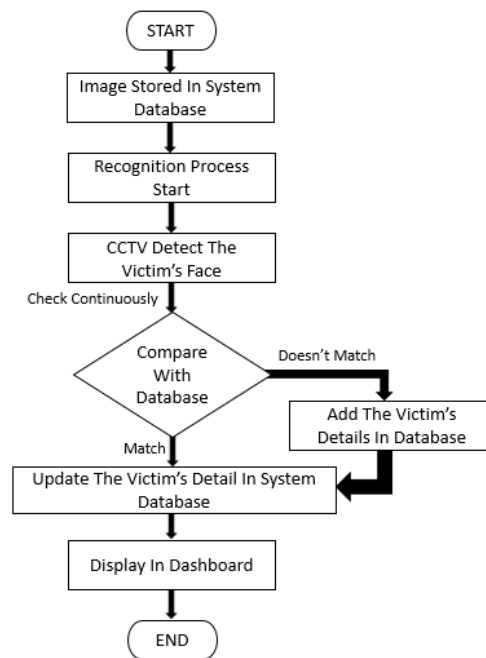


Fig. 6. Flowchart of flood emergency response system with face analytics.

The dashboard website is used to provide analytical data such as the number of people in the shelter, video footage of facial recognition results, and the identities of the people in the evacuation center. This analytical data is used to identify the victims' faces. The dashboard is developed to visualize data [21],[22] for easy to be understand, producing interesting and colorful diagram to distinctive the flood occurrence, maximum flood duration, number of affected flood victims and estimated loss [14].

### V. RESULTS AND DISCUSSION

Based on findings of the FERSFAD system, most researchers and users used for verify or identify the flood victim using their face. The CCTV camera help to capture, analyze and compares pattern based on the victim's facial details [11]. Other than that, the CCTV detects the victim faces to process with other step in detecting and localizes the victims face in images [11], [12]. The process of the capture the victim's face transforms analog into a set of digital information such as database on the victim's facial features. If the victims face match, it verifies the similar faces that belong to the same person. When compare with other biometric technologies such as RFID card, fingerprint scanner and iris templates, the face recognition gives faster detection to verify the flood victim face over an overall victim at the evacuation center [10]. Fig. 7 illustrates the architecture of the FERSFAD system.

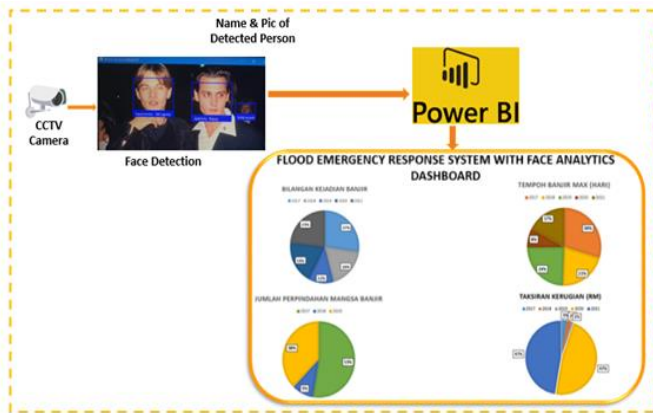


Fig. 7. Flood emergency response system with face analytics architecture.

For instance, there are 230 victims at the evacuation center. Face recognition detects a victim's face using face recognition takes roughly about a second (0.8 second) that is faster compared to fingerprint recognition and took more than two seconds in time (2.3 second). Nonetheless, recognition using an RFID card or fingerprint scanner would take longer time to identify each victim.

The HOG algorithm is included into the system because color data is less important in matching phase for subsequent processing, since the enrolment phase begins by converting the flood victim face image to monochrome in order to detect the victim's face. It evaluates each pixel individually and looks for pixels that are directly connected to (around) the selected pixel. It captures how dark the selected pixel is in comparison to pixels immediately around it. An arrow denotes the magnitude in which the image is darker. The image receives an arrow in

every pixel by repeating this technique for every pixel in the image. These arrows are known as gradients, and they indicate the transition of light to dark throughout the whole image [11-20].

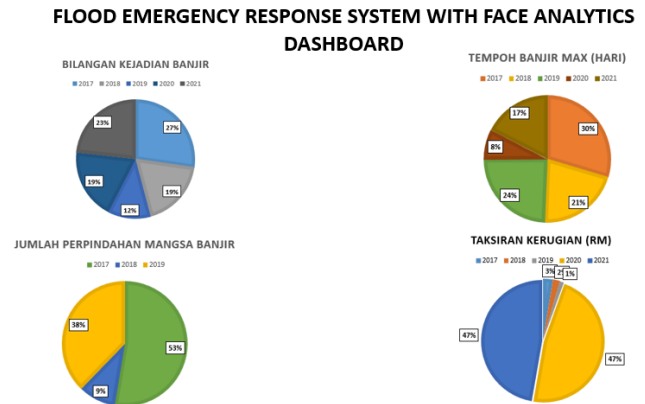


Fig. 8. Dashboard of flood emergency response system with face analytics.

Fig. 8 shows the pie charts in the flood emergency response with face analytics. The pie charts are divided into four parts which are the number of flood occurrence, maximum flood period (Days), the number of displaced flood victims and loss assessment for five years starting from 2017, 2018, 2019, 2020 and 2021. The upper left most section in the pie chart describes about the amount of flood occurrence, which in the highest percentage of the flood occurrence occurred in 2017. Meanwhile, the lowest percentage of the flood occurrence happened in 2019. Thus, in 2017 many flood victims have been affected by the flood disaster.

The upper right most section of the pie charts shows the flood duration in days. To compare data in 2017 and 2019, there are 6% of different between of these two years. The minimum of the flood period (days) in 2020 is with 8%. However, the information about the flood duration has shown a decrement of percentage value due to improving the infrastructure done by the community such as higher dam, enlarging the water channel and making a concrete at the river banks.

The lower left most chart compares for three years of number of affected flood victims which in 2017, 2018 and 2019. In 2017 is the highest percentage of number of affected flood victims, which is 53%. In the second stage in 2019, gives 38 % and the lowest percentage is 9% that is in 2018. The highest number of affected flood victims happened due to no preparation for dealing with flood disaster that the flood victims not aware about the flood disaster.

The lower right most side of the pie charts explains about the estimated loss. Based on the estimated loss information, it is stated that in 2020 and 2021, the amount of loss is 42% and this value is the highest percentage among other values. Even though the duration of floods has been reduced, however, the affected area of flood consists of huge locations in Malaysia since monsoon season, moon's elliptical orbit and the related gravitational pull on Earth. Therefore, many flood victims have been affected tremendously by the flood disaster.

## VI. CONCLUSION

As a conclusion, a flood emergency response system with face analytics has been implemented for flood disaster management system in Malaysia. The flood emergency response system with face analytics has aided evacuation centers in identifying flood victims, strategies to improve the infrastructure and for budget planning. The dashboard displays the actual result of flood victims that gathered from various evacuation centers in Malaysia. The benefit of employing a flood emergency response system with face analytics can aid in data analytics and provides real-time information. Furthermore, the dashboard enables the community to monitor the current situation of flood using their gadgets, phone, tablet, laptop, and other devices. The significant of the dashboard assist victims, police, fire department, and NADMA in preparation for facing the flood disaster. In fact, the flood prediction is estimated and planned to be further investigated in the future.

## ACKNOWLEDGMENT

Thank you to Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka. Authors show appreciation for the funding of this short term grant: PJP/2020/FTMK/PP/S01800 – A Rapid Flood Emergency Response for Prediction Based on Crowd Source Data.

## REFERENCES

- [1] S. Annamalai, P. Caspellanarce, S. Mamat, M. Mat Amin and S. S. Purewal, "Malaysia: Disaster Management Reference Handbook (June 2019)", pp. 1-96. <https://reliefweb.int/report/malaysia/malaysia-disaster-management-reference-handbook-june-2019>.
- [2] H. Abdul Rahman, "Youth and Disaster Management in Malaysia, International Journal of Academic Research in Business and Social Sciences", vol 10, no 16, pp. 367-380, 2020.
- [3] N. Omar Chong and K.H. Kamarudin, "Disaster Risk Management in Malaysia: Issues And Challenges From The Perspective of Agencies" Planning Malaysia Journal, vol. 16, no 5, 2018.
- [4] V. M. Cvetković. "Innovative Solutions for Disaster Early Warning and Alert Systems: A Literary Review", Thematic Conference Proceedings of International Significance, vol 11, no. 2021, 2021. <https://eskup.kpu.edu.rs/dar/article/view/279/203>.
- [5] J. Rajivgandhi, A. Archana, P. Priyanka and C. Shanjana, "Live Coverage of Flood Alert System". International Journal of Engineering Research and Technology, vol 7, no 6, pp. 38-43, 2019.
- [6] K. Schröter, S. Lüdtkke, D. Eggert, H. Kreibich and B. Merz, "Data integration for rapid flood mapping and impact assessment", Budapest University of Technology and Economics, pp. 1 – 6, 2021.
- [7] S. Janizadeh, S. Chandra Pal, A. Saha, I. Chowdhuri, K. Ahmadi, S. Mirzaei, A. Hossein Mosavi and J.P. Tiefenbacher "Mapping the spatial and temporal variability of flood hazard affected by climate and land-use changes in the future, Journal of Environmental Management", vol 298, pp. 113551. <https://doi.org/10.1016/j.jenvman.2021.113551>.
- [8] Q. Sholihah, W. Kuncoro, S. Wahyuni, S. Puni Suwandi, E. Dwi Feditasari " The analysis of the causes of flood disasters and their impacts in the perspective of environmental law", IOP Conf. Series: Earth and Environmental Science 437 (2020) 012056. <https://doi.org/10.1088/1755-1315/437/1/012056>.
- [9] S. Samansiri and T. Fernando and B. Ingirige. "Advanced Technologies for Offering Situational Intelligence in Flood Warning and Response Systems: A Literature Review" Water, vol. 14, no. 2091, 2020. <https://doi.org/10.3390/w14132091>.
- [10] M. Al Rousan and B. Intrigila " A Comparative Analysis of Biometrics Types: Literature Review", Journal of Computer Science, vol.16, no. 12, pp. 1778-1788, 2020. <https://doi.org/10.3844/jcssp.2020.1778.1788>.
- [11] R. D. Hefni Al-Fahsi, A. Patar Jiwandono Pardosi, K. A. Winanta, T. Kirana, O. F. Suryani and I. Ardiyanto, "Laboratory Attendance Dashboard Website Based on Face Recognition System," 2019 International Electronics Symposium (IES), Surabaya, Indonesia, 2019, pp. 19-23. doi: 10.1109/ELECSYM.2019.8901615.
- [12] H. Yang and X. Han, "Face Recognition Attendance System Based on Real-Time Video Processing," in IEEE Access, vol. 8, pp. 159143-159150, 2020. doi: 10.1109/ACCESS.2020.3007205.
- [13] M. Aware, P. Labade, M. Tambe, A. Jagtap, C. Beldar "Attendance Management System using Face-Recognition", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol.7, no. 3, pp. 336-341, 2021. doi: <https://doi.org/10.32628/CSEIT217370>.
- [14] A. Wu, Y. Wang, M. Zhou, X. He, H. Zhang, H. Qu, and D. Zhang, "MultiVision: Designing Analytical Dashboards with Deep Learning Based Recommendation", 2021 IEEE, <https://doi.org/10.48550/arXiv.2107.07823>.
- [15] K. Sharma, D. Anand, M. Sabharwal, P. Kumar Tiwari ,O. Cheikhrouhou and T. Frikha. "A Disaster Management Framework Using Internet of Things-Based Interconnected Devices", Hindawi, Mathematical Problems in Engineering, vol. 2021, Article ID 9916440, pp. 1-21, 2021. <https://doi.org/10.1155/2021/9916440>.
- [16] M. A. A. Baky and M. Islam and S. Paul., "Flood Hazard, Vulnerability and Risk Assessment for Different Land Use Classes Using a Flow Model", Earth Systems and Environment, vol. 4, no. 12, pp. 1-20, 2020. doi:10.1007/s41748-019-00141-w.
- [17] D. Perera, O. Seidou, J. Agnihotri, H. Mehmood and M. Rasmy, "Challenges and Technical Advances in Flood Early Warning Systems (FEWS)", in Flood Impact Mitigation and Resilience Enhancement, Springer, pp. 1-18, 2020. doi: 10.5772/intechopen.93069.
- [18] D. Perera, O. Seidou, J. Agnihotri, M. Rasmy, V. Smakhtin, P. Coulibaly, H. Mehmood, "Flood Early Warning Systems: A Review Of Benefits, Challenges And Prospects", UNU-INWEH Report Series, Issue 8, pp. 1-30, 2019.
- [19] M. Milašinović; D. Prodanović; B. Zindović; B. Stojanović; N. Milivojević, "Control theory-based data assimilation for hydraulic models as a decision support tool for hydropower systems: sequential, multi-metric tuning of the controllers", Journal of Hydroinformatics, vol. 23, no. 3, pp. 500–516, 2021. <https://doi.org/10.2166/hydro.2021.078>.
- [20] R. Parashivamurthy and C. Naveena and Y. H. Sharath Kumar, "SIFT and HOG features for the retrieval of ancient Kannada epigraphs", IET Image Processing, vol. 14, no. 17, pp. 4657-4662, 2020. <https://ietresearch.onlinelibrary.wiley.com/doi/pdfdirect/10.1049/iet-ipr.2020.0715>.
- [21] M. Nadj and A. Maedche and C. Schieder. "The effect of interactive analytical dashboard features on situation awareness and task performance", Decision Support Systems, vol. 135, no. 2020, pp. 1-13. <https://doi.org/10.1016/j.dss.2020.113322>.
- [22] B. Bach, E. Freeman, A.A. Rahman, C. Turkey, S. Khan, "Dashboard Design Patterns," in IEEE Transactions on Visualization and Computer Graphics, vol. 29, no. 1, pp. 342-352, Jan. 2023. doi: 10.1109/TVCG.2022.3209448.

# A Novel Smart Deepfake Video Detection System

Marwa Elpeltagy<sup>1</sup>, Aya Ismail<sup>2\*</sup>, Mervat S. Zaki<sup>3</sup>, Kamal Eldahshan<sup>4</sup>

Systems and Computers Department, Al-Azhar University, Egypt<sup>1</sup>

Mathematics Department, Tanta University, Egypt<sup>2</sup>

Mathematics Department, Al-Azhar University (Girls Branch), Egypt<sup>3</sup>

Mathematics Department, Al-Azhar University, Egypt<sup>4</sup>

**Abstract**—Rapid advancements in deep learning-based technologies have developed several synthetic video and audio generation methods producing incredibly hyper-realistic deepfakes. These deepfakes can be employed to impersonate the identity of a source person in videos by swapping the source's face with the target one. Deepfakes can also be used to clone the voice of a target person utilizing audio samples. Such deepfakes may pose a threat to societies if they are utilized maliciously. Consequently, distinguishing either one or both deepfake visual video frames and cloned voices from genuine ones has become an urgent issue. This work presents a novel smart deepfake video detection system. The video frames and audio are extracted from given videos. Two feature extraction methods are proposed, one for each modality; visual video frames, and audio. The first method is an upgraded XceptionNet model, which is utilized for extracting spatial features from video frames. It produces feature representation for visual video frames. The second one is a modified InceptionResNetV2 model based on the Constant-Q Transform (CQT) method. It is employed to extract deep time-frequency features from the audio modality. It produces feature representation for the audio. The corresponding extracted features of both modalities are fused at a mid-layer to produce a bimodal information-based feature representation for the whole video. These three representation levels are independently fed into the Gated Recurrent Unit (GRU) based attention mechanism helping to learn and extract deep and important temporal information per level. Then, the system checks whether the forgery is only applied to video frames, audio, or both, and produces the final decision about video authenticity. The newly suggested method has been evaluated on the FakeAVCeleb multimodal videos dataset. The experimental results analysis assures the superiority of the new method over the current-state-of-the-art methods.

**Keywords**—Deepfake; deepfake detection; bimodal; XceptionNet; InceptionResNetV2; constant-Q transform; CQT; Gated Recurrent Unit; GRU; video authenticity; deep learning; multimodal

## I. INTRODUCTION

The rapid development of artificial intelligence techniques: autoencoders, Generative Adversarial Networks (GANs), and variational autoencoders facilitated the generation of hyper-realistic fake videos, images, and audio. A deepfake indicates a synthetic image or video AI-generated by swapping an individual's face with another. Applications such as ZAO [1], and DeepFaceLab [2] enable individuals to rapidly generate forged images and videos easily. Recently, a human's voice can be cloned using advanced AI techniques. AI-based audio manipulation is a category of deepfake that clones a human's voice and shows that human saying things that he never said.

Overdub, iSpeech, and VoiceApp are instances of voice cloning open-access platforms that can generate synthesized deepfake sounds that nearly resemble the target human's speech [3]. The work of [4] is an example of these manipulation methods, which involves the creation of highly realistic deepfake videos with a precise lip-sync using a group of AI technologies; FaceSwap, FaceSwap GAN, DeepFaceLab, SV2TTS [5], and Wav2Lip [6].

The majority of deepfake videos are created by cloning sounds, synchronizing lips, and frame-by-frame synthesizing faces. Nevertheless, they lack natural emotions, pauses, and breathing behaviour. Additionally, they suffer from discontinuity and faces' flickering among frames. Deepfake can be misused to impersonate individuals, configure an opinion towards a public figure, and spread falsified news. Therefore, a deepfake detection method is needed to cope with the progress in the deepfake generation process and to distinguish the fakes in video frames, audio, and the whole video including video frames and audio.

This paper introduces a smart deepfake detection method that captures the manipulation in a video (multimodal by nature) on three levels; video frames, audio, and the whole video. It distinguishes whether a given video is a deepfake or not. Two proposed feature extraction methods are employed to extract features from video frames and audio modalities. The first method applied to the visual video frames modality is the XceptionNet with some newly introduced modifications. The Xception network achieved effective results in distinguishing the manipulated videos [7, 8]. The suggested modifications to the Xception network produce useful spatial information of the video frames and improve the deepfake detection method performance. The second method applied to the audio modality is a modified InceptionResNetV2 model based on the CQT method to produce deep time-frequency information of the audio segments and improve the detection method performance. The CQT is a time-frequency analysis method that produces higher time resolution at high-frequency areas and higher frequency resolution at low-frequency areas [9]. Its efficiency has been proven in music signal processing tasks [10], speaker verification systems [11], acoustic scenes and events detection and classification [12], anti-spoofing [13], synthetic speech detection [14, 15], and speech emotion recognition [9]. The corresponding features extracted from the two modalities are fused at a mid-layer to create a bimodal information-based feature representation for the whole video. Finally, the GRU-based attention mechanism is applied to these three levels of representation independently. This assists to learn instructive temporal information for each level and

\*Corresponding Author.

detect deepfake videos. The GRU performs well in tasks of sequence learning and overcomes the gradient vanishing and explosion problems of the standard recurrent neural network [16]. The proficiency of the attention mechanism has been proven in several areas including machine translations, image captioning, question answering, speech recognition [17], and event detection [18]. A comparative study with recent state-of-the-art deepfake detection methods is conducted in terms of accuracy, Area Under Receiver Operating Characteristic (AUROC) curve metric, precision, recall, F1-score, sensitivity, and specificity.

The rest of this work is organized as follows: Section II presents the literature review for deepfake video detection methods. Section III presents the newly proposed method for deepfake video detection. Section IV is dedicated to the experimental results and analysis. The conclusion and future work are presented in Section V.

## II. LITERATURE REVIEW

The progress of AI-based video and voice generation methods raised the ease of creating natural and highly realistic deepfakes that can never be distinguished. Since deepfakes violate security and pose a real threat to society, several researchers have directed their interest to create methods for detecting deepfakes. However, they concentrate on detecting the deepfakes either in video frames or audio modality.

Some of the existing deepfake visual video detection methods spot the manipulation by targeting specific spatial and temporal artifacts that are generated during the fake creation process. Some other detection methods are data-driven that do not target any specific artifacts and distinguish the manipulation by classification [3]. The deepfake visual video detection methods can be categorized into Convolution Neural Network (CNN)-based methods [19, 20, 21, 22], methods that are based on CNN with a temporal network [23, 24, 25, 26, 27], handcrafted feature-based methods [28], and handcrafted feature-based methods with deep networks [29, 30]. This is illustrated in Fig. 1.

The work of [19] detected the deepfakes by exploiting artifacts left by the generation methods when warping the target image to be consistent with the source video. It used four pre-trained CNN models for detecting fake contents; ResNet101, VGG16, ResNet50, and ResNet152. Since deepfake videos suffer from inconsistency among the inter-frames, Hu et al. [20] introduced two branches that are based on CNNs to capture those local and global inconsistencies and then detect deepfakes. Rana and Sung [21] proposed a deep ensemble learning method for detecting deepfake videos. Their method depended on combining several deep base-learners and then training a CNN on these learners to build an ameliorated classifier. In [22], a fine-tuned InceptionResNetV2 model followed by the XGBoost model was employed to capture discrepancies in the spatial domain of fake videos and then individuate deepfakes. The FakeApp creates forged videos that had intra-frame and temporal inconsistencies between frames.

Such inconsistencies were detected using InceptionV3 CNN and long short-term memory (LSTM) models [23]. As AI-generated fake videos lack normal eye blinking, Li et al. [24] introduced the VGG16-LSTM to capture the temporal regularities in the eye blinking process and then distinguish the deepfakes. Most deepfake videos are created frame-by-frame where each forged face is created independently. This causes incoherence in the temporal domain of the face region; discontinuity and flickering. As a result, Zheng et al. [25] introduced a fully temporal convolution network that aimed to learn the temporal discrepancies while removing spatial ones. Then, a temporal transformer encoder followed by a multi-layer perceptron was employed to learn the long-range inconsistencies along the time dimension, and then distinguish the deepfakes. In [26], a 2D CNN-based Spatio-temporal learning model was introduced to learn and capture spatial and temporal inconsistencies of forged videos. This temporal inconsistency was captured from both vertical and horizontal directions over adjacent frames and helped in detecting the fakes. The work of [27] introduced a fine-tuned EfficientNet-b5 model followed by the bidirectional LSTM model and densely connected layer. It aimed to discover the Spatio-temporal inconsistencies in deepfake videos and then distinguish the authenticity of videos. Deepfakes were created by joining the generated face into the source image. This produced errors in facial landmark locations that were detected by estimating the 3D head poses for real and deepfake videos. Then, the estimated difference of head poses was fed into the Support Vector Machine (SVM) for deepfake detection [28]. Khalil et al. [29] proposed a model that employed the local binary patterns descriptor to analyze the texture of real and fake videos. Additionally, a CNN-based enhanced high-resolution network was used to automatically capture informative multi-resolution representations of these videos. Then, the output of both was fed into the capsule network to individuate deepfakes. Ismail et al. [30] introduced a hybrid method in which two feature extraction methods were employed to learn and extract enrich spatial features from the detected face frames of video. These methods were a CNN that was based on the Histogram of Oriented Gradient (HOG) method and the improved XceptionNet. Their outputs were merged to be fed into GRUs sequence to extract the spatiotemporal features and detect the fake videos.

The deepfake audio detection methods can be categorized into handcrafted feature-based methods [31, 32], methods that are based on low-level features with CNN [14, 33, 34, 35], methods that rely on using low-level features with CNN and temporal network [37, 38], and end-to-end deep networks-based methods [39]. This is presented in Fig. 2.

The work of [31] extracted several low-level-features; Constant-Q Cepstral Coefficients (CQCC), Cepstrum, Mel-Frequency Cepstrum Coefficients (MFCC), inverted MFCC, Linear Predictive Cepstral Coefficients (LPCC), and LPCC-residual features. These features were utilized along with the Gaussian Mixture Model (GMM) to detect the forged audio.

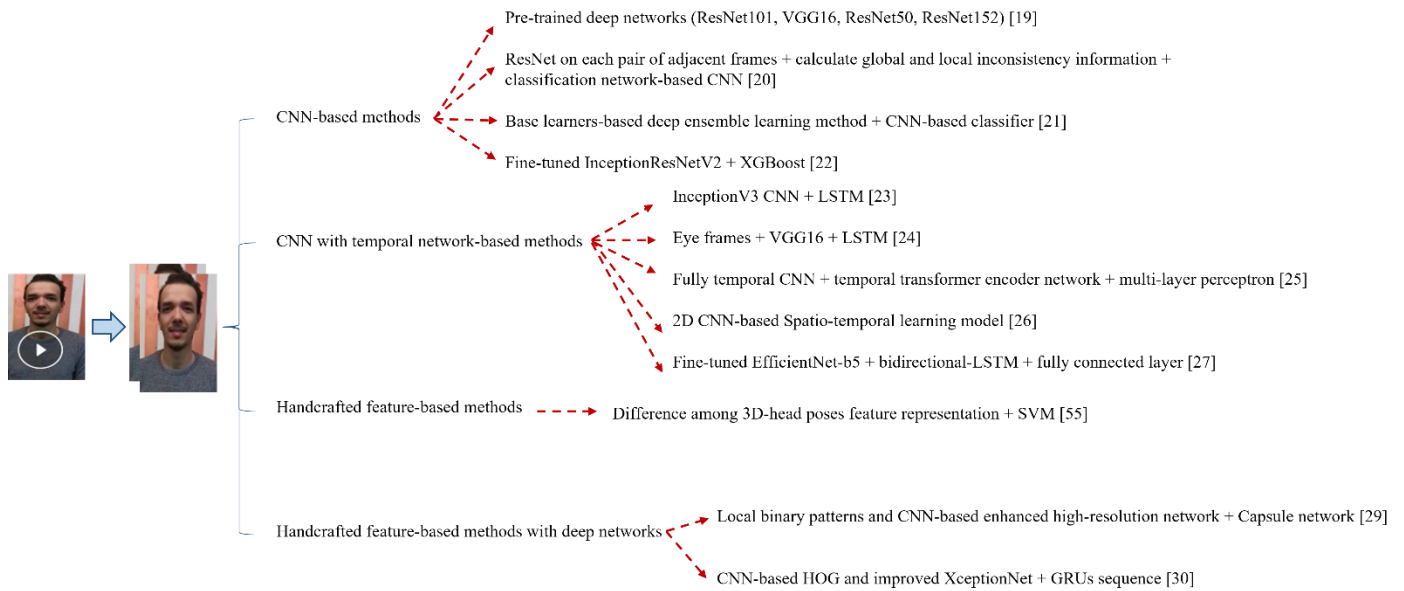


Fig. 1. The deepfake visual video detection methods categorization.

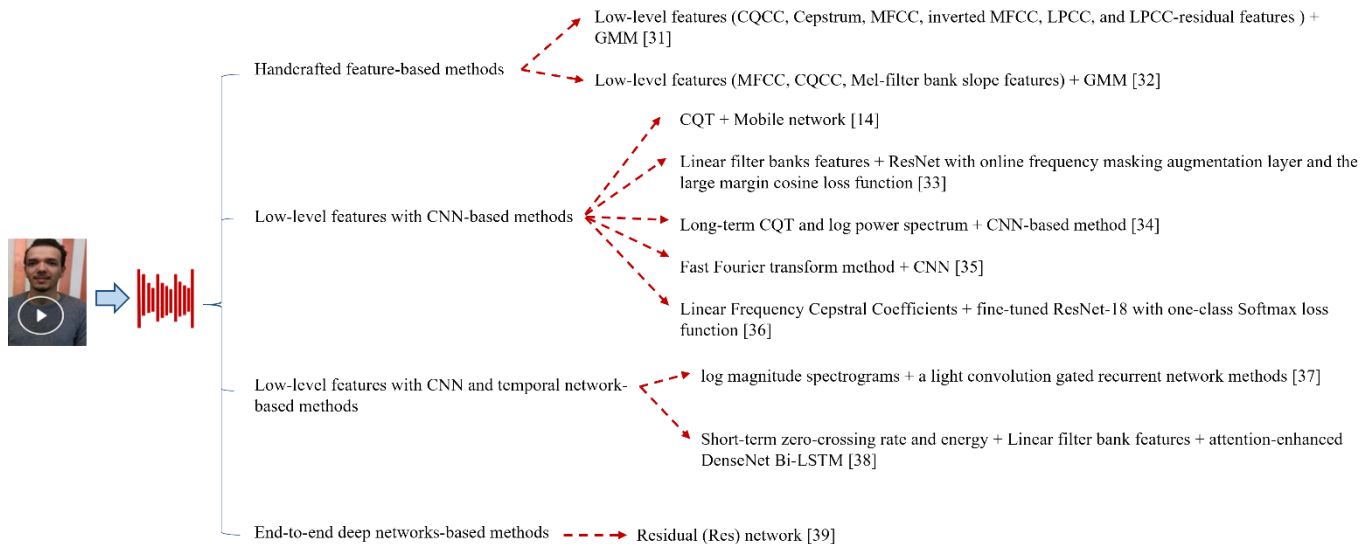


Fig. 2. The deepfake audio detection methods categorization.

In [32], MFCC, CQCC, and Mel-filter bank slope features were employed to train the GMM to capture the vocal tract information and then distinguish the fake audio. Reimao [14] employed the CQT method to convert the audio signals into visual audio representations. These produced representations were fed into the Mobile network model to detect the synthetic speech. The linear filter banks' low-level features were extracted from audios. Then, these features were fed into the ResNet model to produce deep feature representations and detect audio manipulation. In addition, the online frequency masking augmentation layer and the large margin cosine loss function were employed during training the Residual network to learn more robust key feature embeddings [33]. WU et al. [34] employed the long-term CQT and log power spectrum to

extract audios representation. This representation was used as an input to the feature genuinization method. This method learned a transformer with a CNN which was based on genuine speech characteristics. It aimed to maximize the difference between the distribution of genuine and synthetic speeches. Then, the transformed features were utilized with a light CNN model to detect the synthetic speech. In [35], the audio signals were converted into spectrogram images using the Fast Fourier transform method. These images were fed as an input into a CNN to validate audio signal authenticity. The work of [36] utilized Linear Frequency Cepstral Coefficients to convert raw audios into feature vector representations. Then, these representations were fed to a fine-tuned ResNet-18. In addition, a one-class Softmax loss function was proposed to

learn an embedding feature space in which the genuine speech had a compact boundary while the fake data was isolated from the genuine one by a certain margin. In [37], the log magnitude spectrograms were extracted from audio files. Then, a light convolution gated recurrent network was employed on these spectrograms to produce deep features and discriminate the real speech from the spoofed one. The work of [38] employed the short-term zero-crossing rate and energy to select the silent segments from each speech signal. Then, the linear filter bank features were extracted and fed into an attention-enhanced DenseNet Bi-LSTM model to identify audio manipulations. In [39], an end-to-end model which is based on the Residual network was proposed to extract deep features of audio data and then detect the synthetic speech.

Some researchers introduced approaches based on learning from different modalities to detect deepfakes. These approaches, which are often known as deepfake multimodal-video detection methods, can be categorized into CNN-based methods [40, 41, 42, 43], and methods that are based on using CNN and temporal network [44, 45]. This is depicted in Fig. 3.

The work of [40] exploited the perceived emotion cues from speech and face modalities to detect any manipulation in a video. It employed the OpenFace-V1 technique to extract the facial features and the PyAudioAnalysis library to extract the MFCC speech features. Then, the Siamese network-based architecture and the triplet loss were utilized to model the similarity between both modalities within a video and distinguish the fake content. Since any modification of visual video frames or audio modality within a video lead to a loss of lip synchronization, and abnormal lip and facial movements, a multimodal video deepfake detection method was introduced [41]. This method was based on computing the dissimilarity score between visual video and auditory segments. The 3D-Residual network-based architecture was used for extracting visual video features from face segments, and the raw audio segments were converted into MFCC features and then fed into CNN. The contrastive loss was estimated over audio and visual video features for each segment, which forced the real representation of both modalities to be closer than the manipulated one. Additionally, the cross-entropy loss was applied on every single modality to confirm that each one independently learns informative features. The work of [42] presented a multimodal video deepfake detection method based on discovering the defects in manipulated mouth areas via employing genuine audio as a reference. The audios were aligned and clipped into partitions based on phonemes, and Mel-scale spectrograms were extracted and used as audio features. The mouth frames were extracted from videos based on facial landmarks using the `dlib` python library. Then, each mouth frame with a particular phoneme interval was matched to a fixed-length audio partition to produce auditory-visual video pairs. After that a CNN architecture was trained on these pairs to capture the synchronization degree between lip movements and speech by measuring the similarity score of

auditory-visual video pairs. Zhou and Lim [43] employed the asynchrony property between fake visual video, especially mouth movements, and speech to detect any modification within a video. The Multi-Task CNN (MTCNN) was utilized for detecting the face from video frames and the Residual (2+1)D-18 network was applied for extracting visual feature representations of these frames. For audio, a simple 1D convolution network was utilized for extracting 1D waveform signal feature representation. In addition, a sync-stream was built by applying central connections to visual video and audio network feature representations between low-level features; spatial and temporal information, to higher-level semantic representations. At each layer, the representations of visual video and auditory modalities were fused with the current layer of sync-stream. The output of this was utilized as an input to the fusion at the next layer. This helped in modelling the synchronization patterns of both modalities and distinguishing the deepfakes. Based on the observation that machines cannot recreate human emotions naturally in manipulated videos, Gino [44] introduced a deepfake detection method depending on exploiting emotion features from visual video and audio modalities. The low-level descriptors (LLDs) were extracted from raw audio segments using the OpenSmile toolkit and passed to the LSTM architecture to extract emotional features of speech. In addition, the face frames were detected from videos using the BlazeFace tool and then passed into 3D-CNN architecture to extract visual emotional features. After that, two approaches were followed in the final deepfake detection phase. In the first one, the visual and auditory emotional features were combined horizontally. Then, these features were fed either into the LSTM network or into Lazypredict models. In the second, the average between the prediction scores returned by training the LSTM and Lazypredict models on the visual video and auditory modalities separately was computed. The work of [45] detected the fake content in videos by extracting visual video and auditory emotional features and passing them to a deep network. The OpenFace-V2 toolkit was employed for extracting 31 visual features related to the intensity of facial muscle actions, eye gaze, and head position. The `python_speech_features` library was used to extract 13 MFCC features and their respective derivatives; delta MFCC, and delta-delta MFCC, as audio features. The visual video and auditory features were normalized and concatenated to be passed into CNN blocks that were followed by two Bi-LSTM networks and dense layers for deepfake detection.

A few deepfake detection methods are concerned with multimodal videos. However, they do not consider whether a video is manipulated only on the visual video frames level, audio level, or bimodal level which combines visual frames and audio. Consequently, this paper introduces a novel smart deepfake video detection system that can check whether the manipulation is just applied to video frames, audio, or both. It then produces the final decision for detecting the deepfakes on these three levels: visual frames, audio, and the whole video.

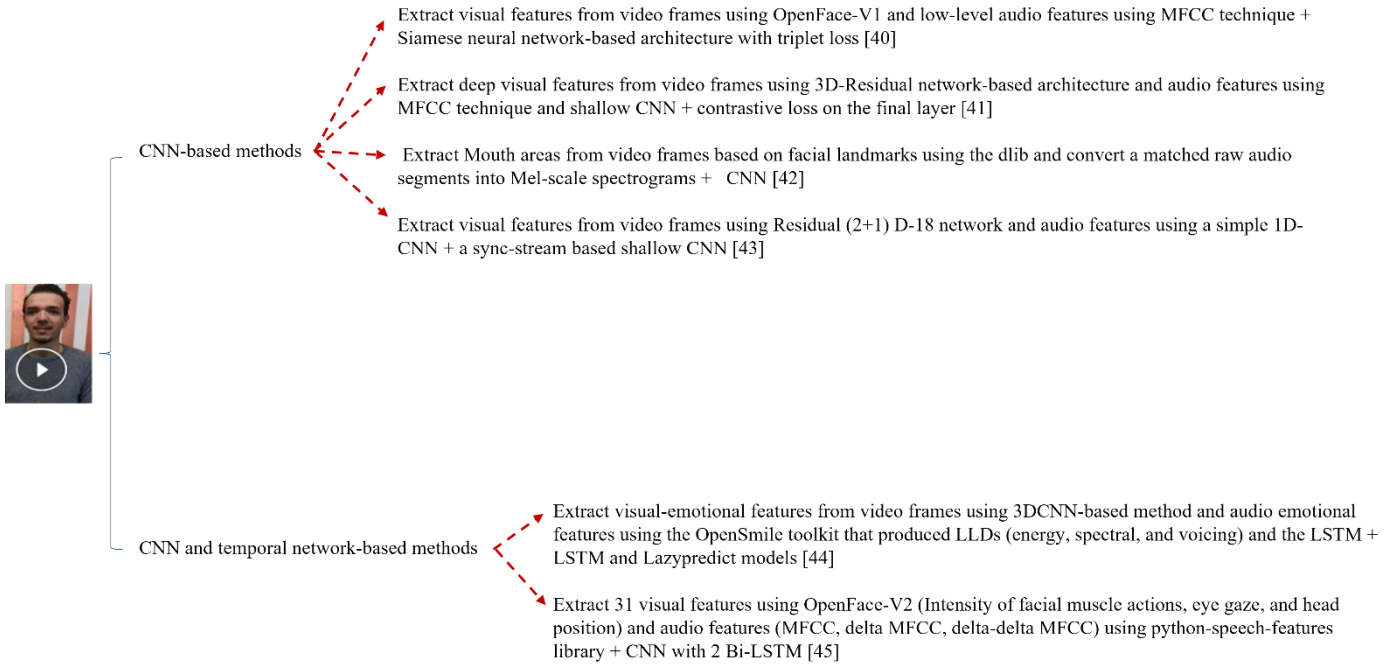


Fig. 3. The deepfake multimodal-video detection methods categorization.

### III. PROPOSED METHOD

The suggested deepfake video detection method consists of three base stages: pre-processing, feature extraction using unimodal and bimodal information, and classification. These stages are shown in Fig. 4 and each one will be described hereafter.

#### A. Pre-processing

The visual video and audio modalities are pre-processed individually. The face frames are extracted from videos and saved separately. They are rescaled to the size  $224 \times 224$ , and their pixel values are normalized into  $[-1,1]$ . These pre-processed face frames will become an input to the next stage for learning and extracting deep visual video features. The raw audio files are extracted from videos and stored separately in a wave format. Then, the audio files are segmented. The CQT method is applied to every audio segment to produce a time-frequency representation of these segments. The CQT method is used for transforming audio signals from the time domain to the time-frequency domain. In CQT, frequency bins are geometrically spaced and ratios between centre frequencies and bandwidths, which are called Q-factors, of all bins are equal [47, 12]. The CQT of a discrete audio signal  $x(n)$  in the time domain is computed by the following formula [10, 46, 47, 48]:

$$X(m, n) = \sum_{k=n-\lfloor N_m/2 \rfloor}^{n+\lfloor N_m/2 \rfloor} x(k) a_m^* \left( k - n + \frac{N_m}{2} \right) \quad (1)$$

where  $m = 1, 2, \dots, M$  represents the  $m^{\text{th}}$  index of frequency bin,  $\lfloor \cdot \rfloor$  denotes the floor function, and  $x(k)$  represents the  $k^{\text{th}}$  sample of a speech time-domain frame. The symbol  $N_m = \frac{f_{sr}}{f_m} Q \in \mathbb{R}$  indicates window lengths,  $f_{sr}$  represents the sampling rate frequency, and  $f_m = f_1 2^{\frac{m-1}{b}}$

indicates the centre frequency of the  $m^{\text{th}}$  bin. The symbol  $f_1$  denotes the centre frequency of the lowest bin,  $b$  refers to the bins number per octave and practically it determines a trade-off between time and frequency resolution. The factor  $Q = \frac{f_m}{f_{m+1} - f_m} = \frac{1}{2^{\frac{1}{b}} - 1}$  produces a constant frequency to resolution ratio for each bin. The term  $a_m^*(n)$  represents the complex conjugate of the complex-valued time-frequency atoms  $a_m(n)$  which is defined as follows:

$$a_m(n) = \frac{1}{C} \omega \left( \frac{n}{N_m} \right) e^{i(2\pi n \frac{f_m}{f_{sr}} + \Phi_m)} \quad (2)$$

where  $\omega(t)$  denotes a window function; Hann (Wang et al. 2019) [49], which is sampled at points specified by  $\frac{n}{N_m}$ . It is zero when  $t$  does not belong to  $[0,1]$ . The  $C = \sum_{l=-\lfloor N_m/2 \rfloor}^{\lfloor N_m/2 \rfloor} \omega \left( \frac{l + N_m/2}{N_m} \right)$  represents a scaling factor, and  $\Phi_m$  represents a phase offset.

The CQT computations are implemented using the librosa python library. The audio files are resampled to 22,050 Hz. A frequency bins number of 84 with 12 bins per octave, a hop length of 128 samples, and a minimum frequency value of approximately 65 Hz are used during the CQT calculations. In addition, the Hann window function is applied. The output of the CQT is then transformed into a log scale; decibels, to cope with the wide range of sound intensity. This produces a decibel-scaled spectrogram that has the shape  $T \times 84$  per audio segment where  $T$  relies on the audio file duration. The duration of audio files adopted here is three seconds and accordingly,  $T$  is equal to 65. The spectrograms are normalized into the range  $[-1,1]$  and then reshaped to  $(65,84,1)$  as three-channel images. They will become an input to the next stage to learn and extract deep auditory features.



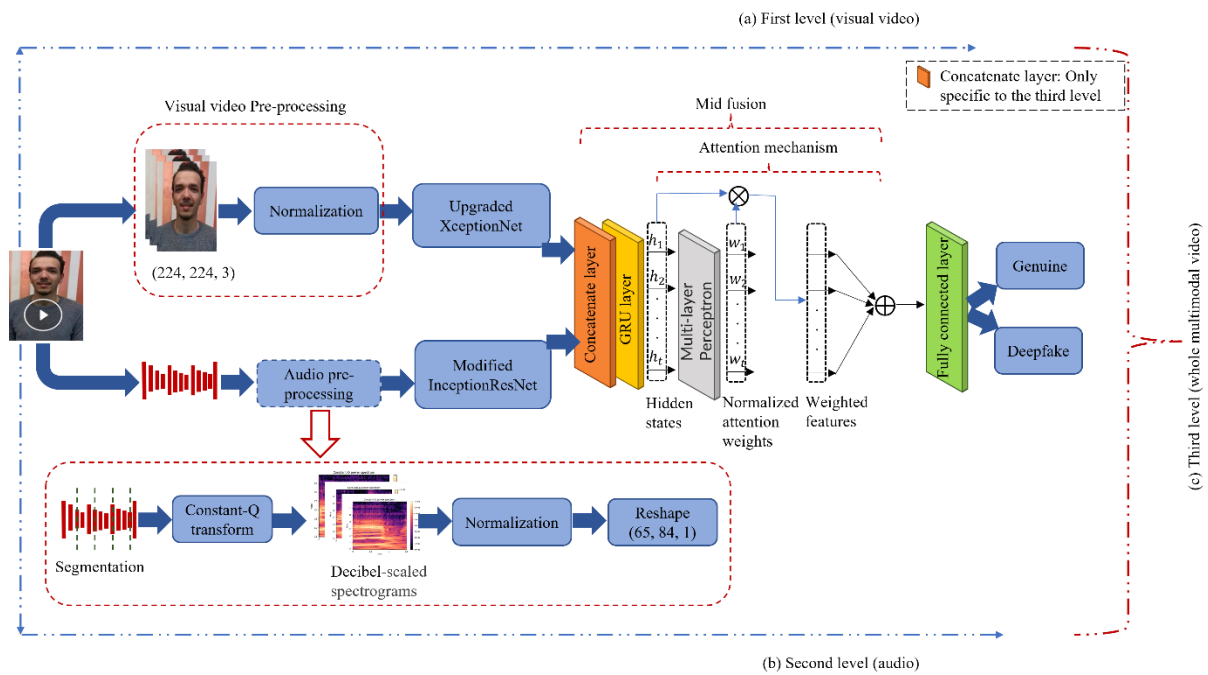


Fig. 4. The proposed smart deepfake video detection system architecture.

### B. Feature Extraction using Unimodal and Bimodal Information

In this stage, the problem of deepfake detection is handled based on proposing two feature extraction methods for visual video and audio modalities. An upgraded XceptionNet is suggested to extract instructive deep spatial features from pre-processed face frames of videos. It outputs a visual feature representation of the unimodality; video frames. A modified InceptionResNetV2 is suggested to apply on the CQT spectrograms representing audio files to extract deep time-frequency features from audios. It produces a feature representation of the unimodality; audio. Then, the corresponding extracted feature representations from these modalities are first fused. This outputs a feature vector representation of the whole video using bimodal information. After that, these various resultant representations are independently passed into the GRU-based attention mechanism. This helps to learn the significant temporal information from the sequential feature representation per video on three levels: visual video frames, audio, or the whole video. Finally, a fully connected layer is applied to produce the final prediction about video authenticity. These components are explained in detail in the following subsections.

1) *Visual video frames features:* The processed face frames of videos with the shape  $(h \times w \times 3)$  are received as an input to the proposed upgraded Xception network where  $h=224$ ,  $w=224$ , and 3 denote the height, width, and RGB channels per frame. The Xception original architecture consists of 36 convolutional layers divided into 14 modules. All modules have shortcut residual connections around them except for the first and last ones. The Xception comprises depth-wise separable convolution layers, which reduce the cost of convolution operation dramatically [50, 51]. The proposed

upgraded Xception network architecture is depicted in Fig. 5. The original XceptionNet is upgraded by first injecting seven layers before the last rectified linear unit (ReLU) activation layer of the last module. These seven layers are convolution with 1536 filters, batch normalization, ReLU activation, convolution with 1024 filters, batch normalization, ReLU activation, convolution with 1024 filters, and batch normalization. The convolution layers produce more informative and exclusive feature maps that help to differentiate between real and fake visual videos. The batch normalization layers, which standardize the input, have the effect of drastically speeding up the training and improving the model's performance by providing a modest regularization. The ReLU activation layers, which give a value of zero for all negative input feature values, add a nonlinear property to the model allowing it to understand and learn complex structures in data. Then, the dropout layer that randomly drops out units with a rate of 0.2 is injected between the last ReLU activation and the global average pooling layers to prevent overfitting and boost the model's generalization. After that two layers are injected after the global average pooling layer; the fully connected layer with 1024 units and ReLU activation function, and the dropout layer with a rate of 0.5. After applying the upgraded XceptionNet to the face frames of videos, the output becomes a vector representation of 1024 features per frame. The suggested modifications to the Xception network attempt to generate an instructive spatial hierarchical representation of frames. This helps to improve the deepfake detection method performance in real-world circumstances; number equations consecutively.

2) *Audio features:* The CQT spectrograms of audio files with the shape  $(65, 84, 1)$  per segment are received as an input

to the proposed modified InceptionResNetV2. The InceptionResNetV2 original architecture is built by joining the inception blocks and the skip connections. Each InceptionResNet block contains convolutions of different-sized filters that are combined by skip connections. These skip connections prevent the degradation problem that occurred via deep structures and reduce the time of training [52].

The proposed modified InceptionResNetV2 architecture is depicted in Fig. 6. The original InceptionResNetV2 is modified first by decreasing the repeating times' number of Inception ResNet blocks; A, B, and C, from 5, 10, and 5 to 4, 7, and 3, respectively. Then, some layers are injected after the last InceptionResNet block C and before the global average pooling layer. These layers are convolution with 512 filters on

a kernel size of  $1 \times 1$ , batch normalization, ReLU activation, a couple of convolutions with 1024 filters on a kernel size of  $1 \times 1$  where each one is followed by batch normalization and ReLU activation, and a dropout with a rate of 0.2. After that a fully connected layer with 1024 units and ReLU activation function is injected between global average pooling and dropout layers. In addition, filter units, kernel size, and stride for some layers are altered as shown in Fig. 6. After applying the modified InceptionResNetV2 to audio files segments, the output becomes a vector representation of 1024 auditory features per segment. The proposed modifications to the InceptionResNetV2 aim to generate an informative deep time-frequency representation of audio segments. This aids to enhance the performance of the proposed deepfake detection method.

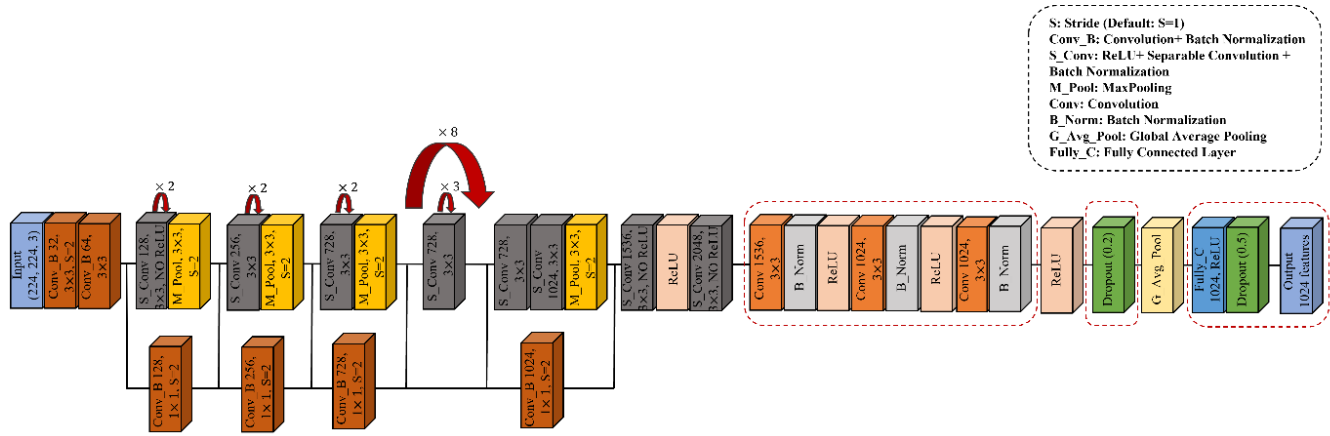


Fig. 5. The proposed upgraded xception network architecture.

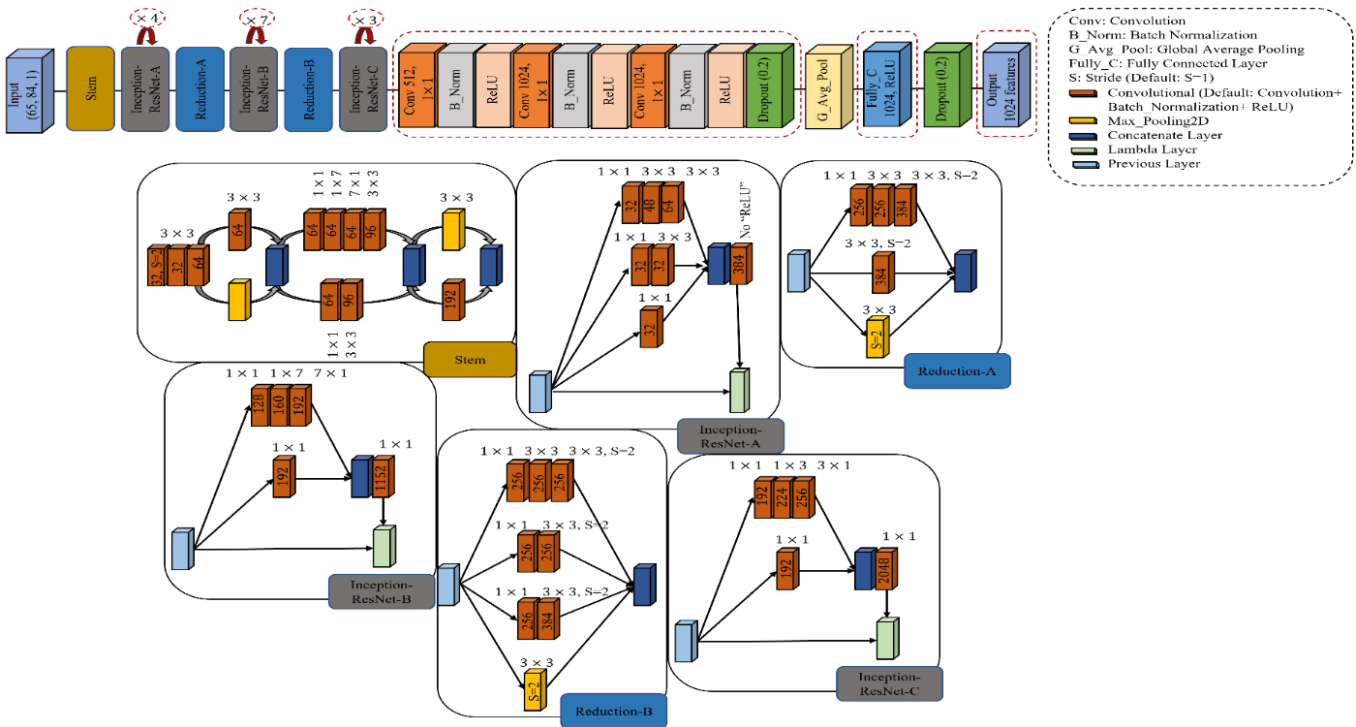


Fig. 6. The proposed modified inceptionResNetV2 architecture.

3) *Bimodal information-based video features*: The deep extracted features from visual video frames and audio modalities using the above-mentioned unimodality-based feature extraction methods are mid-fused at a concatenate layer. This produces a feature vector representation for the whole video, which is based on bimodal information.

4) *Temporal information extraction-based attention mechanism*: Most deepfake videos are generated based on synthesizing faces frame-by-frame, cloning voices, and synchronizing lips. They suffer from flickering and discontinuity of the face frames and lack of normal emotions, breathing, pauses, and the pace at which the target subject speaks among audio segments. As a result, the GRU-based attention mechanism is applied to the three levels of the extracted features independently; visual video frames, audio, and the whole video. This aims to capture the instructive temporal information that helps to differentiate real videos from fake ones.

The GRU architecture is composed of two gates; update (upd) and reset (res), that modulate the information flow from the previous time step to the current step. At each time step  $t$ , the update gate decides the amount of previous information that should be retained, and the reset gate determines the amount of information that needs to be forgotten [53]. The GRU hidden state  $h$  at the time  $t$  is defined by the following formulae [54]:

$$\text{upd}_t = S(W_{\text{upd}}x_t + U_{\text{upd}}h_{t-1}) \quad (3)$$

$$\text{res}_t = S(W_{\text{res}}x_t + U_{\text{res}}h_{t-1}) \quad (4)$$

$$\hat{h}_t = \tanh(W_h x_t + \text{res}_t \circ U_h h_{t-1}) \quad (5)$$

$$h_t = (1 - \text{upd}_t) \circ \hat{h}_t + \text{upd}_t \circ h_{t-1} \quad (6)$$

where  $x$  refers to the input, and  $W$  and  $U$  represent the weight matrices. The symbol  $S(\cdot)$  represents the sigmoid function,  $\tanh(\cdot)$  represents the Hyperbolic Tangent,  $\circ$  denotes the Hadamard product, and  $\hat{h}_t$  denotes the candidate hidden state. As can be seen in Fig. 4, a single GRU is applied to the above-mentioned feature representations on the three levels. It produced a matrix of hidden state vectors at each time step  $t$ , which represents the learned temporal information per visual video, audio, or the whole video. The hidden state vector is defined as follows:

$$H = [h_1, h_2, \dots, h_t] \quad (7)$$

The attention mechanism uses the weights to concentrate on the important features from the input sequence  $H$ . It is defined by the following equations [17, 55]:

$$u_t = \tanh(W_h x_t + b) \quad (8)$$

$$\alpha_t = \text{softmax}(u_t) \quad (9)$$

$$c_t = \alpha_t h_t \quad (10)$$

$$v = \sum_t c_t \quad (11)$$

where  $u_t$  is a result of feeding a hidden vector  $h_t$  into a single-layer Multi-Layer Perceptron (MLP) with the tanh

activation function.  $W$  represents the weight matrix, and  $b$  refers to the bias term. The symbol  $\alpha_t$  represents the normalized attention weights that are produced by applying the softmax layer to  $u_t$ .  $v$  is a video representation that is formed by summing hidden vectors  $h_t$  weighted by attention weights  $\alpha_t$ .

### C. Classification

After the instructive temporal features are produced from the GRU-based attention mechanism, a fully connected layer is used as an output layer with two classes. Softmax function is used to decide deepfake videos from real ones. The Softmax formula is defined as follows:

$$\text{Softmax}(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}} \quad (12)$$

where  $y_i$  denotes the values resulting from the output layer neurons.

### D. Dataset

The proposed method has been evaluated on the FakeAVCeleb multimodal videos dataset. This dataset consisted of 490 celebrity genuine videos that were selected from the VoxCeleb2 dataset based on various ethnic groups, gender, and age. Its genuine videos are face-centered and cropped. The fake videos of the FakeAVCeleb dataset were generated using DeepFaceLab, Faceswap, and FSGAN, while fake audios were generated using a real-time voice cloning tool (SV2TTS). Additionally, the Wav2Lip was applied to the deepfake videos to re-enact these videos based on the cloned audios. Thus, the FakeAVCeleb dataset had more realistic deepfakes. The FakeAVCeleb was divided into four groups; genuine visual videos with genuine audios, genuine visual videos with deepfake audios, deepfake visual videos with genuine audios, and deepfake visual videos with deepfake audios [4].

To evaluate the proposed method, 1215 genuine and deepfake videos of the FakeAVCeleb dataset are employed. These videos are divided into three subsets: training, validation, and testing.

## IV. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed deepfake video detection method is evaluated by the FakeAVCeleb dataset. Its performance is assessed using the following evaluation metrics [56]:

$$\text{precision} = \frac{\text{True\_Positives}}{\text{True\_Positives} + \text{False\_Positives}} \quad (13)$$

$$\text{sensitivity} = \text{recall} = \frac{\text{True\_Positives}}{\text{True\_Positives} + \text{False\_Negatives}} \quad (14)$$

$$F_1 - \text{score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (15)$$

$$\text{accuracy} = \frac{\text{True\_Positives} + \text{True\_Negatives}}{\text{True\_Positives} + \text{True\_Negatives} + \text{False\_Negatives} + \text{False\_Positives}} \quad (16)$$

$$\text{specificity} = \frac{\text{True\_Negatives}}{\text{True\_Negatives} + \text{False\_Positives}} \quad (17)$$

$$\text{AUROC} = \int_0^1 \text{sensitivity}((1 - \text{Specificity})^{-1}(x)) dx = p(x_2 > x_1) \quad (18)$$

where True\_Positives denotes deepfake samples' number that is correctly predicted. The False\_Positives represents genuine samples' number that is incorrectly predicted. The False\_Negatives denotes deepfake samples' number that is incorrectly predicted. The True\_Negatives refers to genuine samples' number that is correctly predicted. The symbol  $x_2$  represents the predicted deepfake samples and  $x_1$  denotes the predicted genuine samples. The higher the AUROC curve metric, the better the fake video detection method's performance at individuating the deepfake videos from the genuine ones.

The following three experiments are applied to the FakeAVCeleb dataset:

Experiment 1: This experiment represents applying the proposed method to the FakeAVCeleb videos dataset for two levels; visual video frames and audio. The visual video frames and audio modalities are trained end-to-end separately. Thus, a single GRU-based attention mechanism with 1024 units is independently applied to the visual video features that are extracted using the proposed upgraded XceptionNet and the audio features that are extracted using the proposed CQT based modified InceptionResNetV2. This learns the instructive temporal features for each unimodality. The visual video modality is trained for 32 epochs using the stochastic gradient descent (SGD) optimizer [57] with a learning rate of  $2e^{-3}$  which is decayed by  $4e^{-10}$ , and a momentum of 0.9. The audio modality is trained for 27 epochs using the adaptive moment (Adam) optimizer [58] with a learning rate of  $e^{-3}$ . The batch size is 32. Then, the predictions are produced per modality. The performance of visual video and audio on the FakeAVCeleb dataset is shown in Table I and Table II, respectively. The proposed upgraded XceptionNet with a single GRU-based attention mechanism for the visual video modality has achieved 98.51% accuracy and 98.45% AUROC outperforming recent state-of-the-art methods by a large margin. Additionally, the proposed CQT based modified

InceptionResNetV2 with a single GRU-based attention mechanism for the audio modality has achieved 97.52% accuracy and 97.62% AUROC outstanding other state-of-the-art methods by a large margin.

Experiment 2: In this experiment, the prediction results of visual video frames and audio modalities from experiment 1 are employed to produce the prediction for the whole video. Thus, the whole multimodal-video prediction is decided to be genuine if both modalities are predicted as genuine, otherwise, it's deepfake. Experiment 2 performance for multimodal video deepfake detection is recorded in Table III. It has yielded 96.04% accuracy and 95.49% AUROC.

Experiment 3: This experiment represents applying the proposed method to the FakeAVCeleb videos dataset for the third level; whole multimodal video. As the FakeAVCeleb dataset is distributed into four groups: genuine visual videos and audios, genuine visual videos with fake audios, fake visual videos with genuine audios, fake visual videos and audios, the whole video label ( $y_i$ ) is considered genuine if the label of both visual video ( $y_{iv}$ ) and audio ( $y_{ia}$ ) modalities are genuine, otherwise, it's fake. This can be defined as follows:

$$y_i = \begin{cases} 0, & \text{if } y_{iv} = y_{ia} = 0 \\ 1, & \text{otherwise} \end{cases} \quad (19)$$

The single GRU-based attention mechanism with 3572 units is applied to the bimodal information-based video features. This helps to learn the instructive temporal features for the whole multimodal video. The details of GRU-based attention mechanism layers that are applied on top of bimodal information-based video features are described in Table IV. The proposed method is trained for 24 epochs using the SGD optimizer with a learning rate of  $2e^{-3}$  and a decay factor of  $4e^{-10}$ , and a momentum of 0.9. This is employed to update the weight parameters and is aimed to minimize the difference between actual and predicted labels. The batch size is set to 64.

TABLE I. THE PERFORMANCE OF THE UPGRADED XCEPTIONNET METHOD WITH SINGLE GRU-BASED ATTENTION MECHANISM FOR DETECTING THE DEEPFAKE VISUAL VIDEO UNIMODALITY COMPARED TO RECENT STATE-OF-THE-ART METHODS ON THE FAKEAVCELEB DATASET

Model	Unimodality	
	Visual video	
	Accuracy	AUCROC
Experiment 1 (The proposed upgraded XceptionNet with GRU-based attention mechanism for the first level)	98.51%	98.45%
VGG16 [60]	81.03%	81.04%
Xception [7]	73.06%	73.07%

TABLE II. THE PERFORMANCE OF THE MODIFIED INCEPTIONRESNETV2 METHOD WITH SINGLE GRU-BASED ATTENTION MECHANISM FOR DETECTING THE DEEPFAKE AUDIO UNIMODALITY COMPARED TO RECENT STATE-OF-THE-ART METHODS ON THE FAKEAVCELEB DATASET

Model	Unimodality	
	Audio	
	Accuracy	AUCROC
Experiment 1 (The proposed CQT based modified InceptionResNetV2 with GRU-based attention mechanism for the second level)	97.52%	97.62%
Mel-frequency cepstrum (MFC)+ VGG16 [60]	67.14%	67.13%
MFC+ Xception [60]	76.26%	76.25%

(CQT [61] + MobileNet) [14]	82.67%	82.38%
-----------------------------	--------	--------

TABLE III. THE PERFORMANCE OF THE PROPOSED METHOD FOR DETECTING WHOLE MULTIMODAL VIDEO DEEPFAKES COMPARED TO RECENT STATE-OF-THE-ART METHODS ON THE FAKEAVCELEB DATASET

Model	Bimodal	
	Visual video and audio	
	Accuracy	AUCROC
<b>Experiment 2</b>	<b>96.04%</b>	<b>95.49%</b>
Experiment 3 (The proposed method for the third level: whole multimodal video)	97.52%	97.21%
Ensemble Soft/ hard voting based VGG16 [60]	78.04%	78.05%
Two CNN blocks (one per modality) [60]	67.4%	67.2%
Xception [7]	43.94%	43.73%

TABLE IV. THE GRU-BASED ATTENTION MECHANISM LAYERS DETAILS

Layer (type)	Output shape	Parameters number
main_input (Input Layer)	[(None, 8, 4096)]	0
gru (GRU)	(None, 8, 3572)	82191720
attention (attention)	(None, 3572)	3580
Total parameters: 82,195,300 Trainable parameters: 82,202,446 Non-trainable parameters: 0		

The cross-entropy loss (l) function is utilized to measure the efficiency of the suggested deepfake video detection method on three levels: video frames, audio, and the whole video. Its formula [59] is defined as follows:

$$l = -\frac{1}{M} \sum_{k=1}^M (y_k \log(p_k) + (1 - y_k) \log(1 - p_k)) \quad (20)$$

where M refers to the number of visual videos, audios, or whole videos. The  $y_k$  and  $p_k$  denote the actual label and predicted probability corresponding to the  $k^{\text{th}}$  video. It can be seen in Table III that the proposed method, which represents experiment 3, for whole multimodal video deepfake detection has achieved 97.52% accuracy and 97.21% AUROC. Its performance exceeds that of experiment 2 because experiment 2 is unable to learn intercorrelations between different modalities. Additionally, it outperforms recent state-of-the-art methods by an average growth of 34.4% accuracy and 34.2% AUROC as can be seen in Table III.

The experiments are carried out using an OMEN HP laptop with a 16-gigabyte Intel (R) Core (TM) i7-9750H CPU, a 6-gigabyte RTX 2060 GPU, and Windows 11. The proposed method is implemented using the Python programming language. Python libraries such as Keras, OpenCV, Random, Tensorflow, Numpy, OS, and Librosa are used during the implementation.

The accuracy and loss curves of the proposed method on the training and validation subsets of the FakeAVCeleb dataset for the three levels; visual video frames, audio, and whole multimodal videos, are shown in Fig. 7. Additionally, the proposed method confusion matrix for deepfake video detection on the three levels is depicted in Fig. 8. Furthermore, Fig. 9 shows the receiver operating characteristic (ROC) curve

and the AUROC curve of the proposed method performance. As shown in Fig. 9, the ROC curve is extremely close to the top left ensuring the high performance of the proposed method.

Fig. 10 provides a comparison of the proposed method with contemporary state-of-the-art methods using evaluation metrics. As shown in Fig. 10, the proposed method has yielded better performance in comparison to the other methods on the three levels. It has a precision of 96.91%, recall of 100%, F1-score of 98.43%, and specificity of 97.22% for detecting visual videos. Additionally, it has a precision of 100%, recall of 95.10%, F1-score of 97.49%, and specificity of 100% for detecting audios. Further, it has a precision of 98.43%, recall of 97.66%, F1-score of 98.04%, and specificity of 97.30% for detecting whole multimodal videos.

It can be concluded that the proposed upgraded XceptionNet generated a useful spatial hierarchical representation of faces, which contributed to distinguishing between genuine and fake videos. As well, the proposed CQT-based modified InceptionResNetV2 produced a valuable deep time-frequency representation of audio. This assisted to reveal deepfake videos and improved the detection method's effectiveness. Moreover, a concatenate layer that is applied to the features extracted from visual video and audio modalities produced an informative bimodal representation of videos. In addition, the GRU-based attention mechanism, which is applied to the visual video, audio, and bimodal features, assisted in capturing the most important temporal information of videos. This in turn helped to detect the deepfakes. Furthermore, it can be inferred that correlating features from different modalities can improve the chances of achieving accurate deepfake video detection.

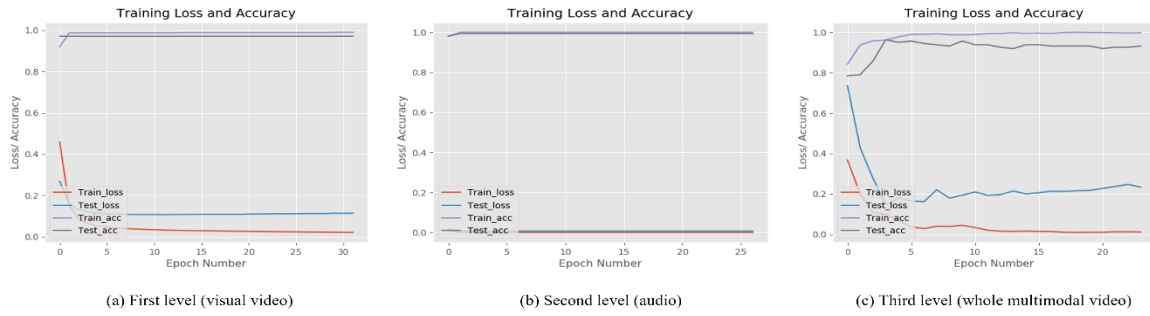


Fig. 7. The accuracy and loss curves of the proposed deepfake video detection method on training and validation sets.

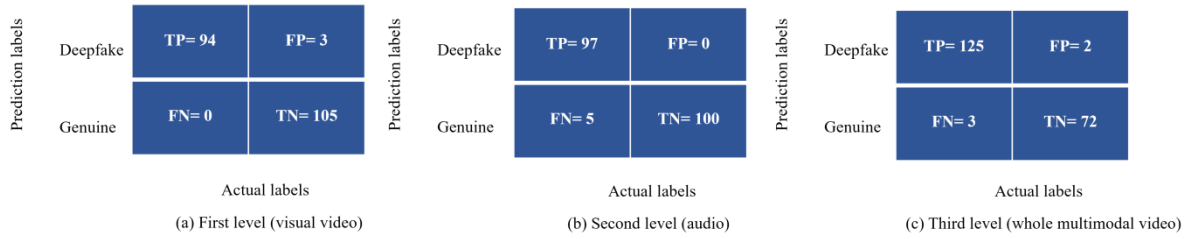


Fig. 8. The confusion matrix visualization of the proposed deepfake video detection method.

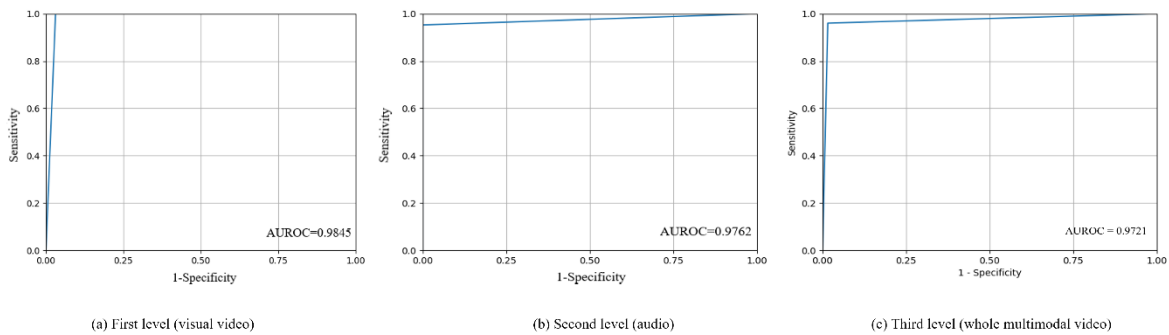


Fig. 9. The ROC curve and the AUROC curve of the proposed deepfake video detection method performance.

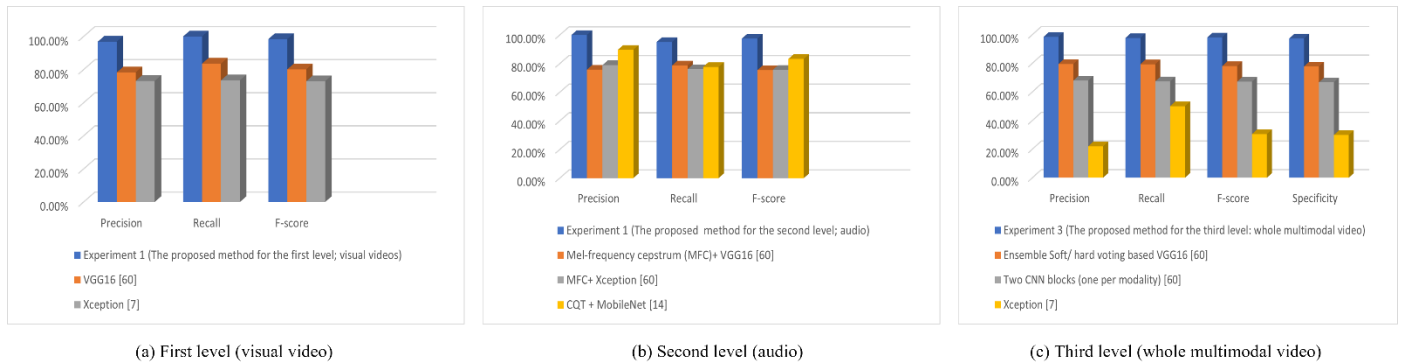


Fig. 10. The evaluation metrics of the proposed deepfake video detection method compared to recent state-of-the-art methods on the FakeAVCeleb dataset.

## V. CONCLUSION AND FUTURE WORK

A newly smart system for detecting video deepfakes has been presented. Two methods were proposed to extract features from visual video frames and audio modalities, respectively. These methods produced useful spatial information for visual video and valuable time-frequency information for audio, which improved the performance of the deepfake detection

method. In addition, the feature representations of both modalities were passed into a mid-layer to produce an informative bimodal representation per video. It proved that using bimodal information boosts learning during training compared to the method that ignores intercorrelation between modalities. The GRU-based attention mechanism was then applied to the different feature representations to extract the most significant temporal information and detect the

deepfakes. The proposed method has been evaluated on the FakeAVCeleb multimodal videos dataset. It achieved 98.51% accuracy, 98.45% AUROC, 96.91% precision, 100% recall, 98.43% F1-score, and 100% sensitivity on the first level; visual videos. Additionally, it yielded 97.52% accuracy, 97.62% AUROC, 100% precision, 95.10% recall, 97.49% F1-score, and 95.10% sensitivity on the second level; audios. Moreover, it attained 97.52% accuracy, 97.21% AUROC, 98.43% precision, 97.66% recall, 98.04% F1-score, 97.66% sensitivity, and 97.30% specificity on the third level; whole multimodal videos. Consequently, the proposed method outperformed the current state-of-the-art methods by a large margin.

In the future, several optimization algorithms can be employed to enhance the performance of the proposed deepfake video detection method. Furthermore, a huge multimodal video dataset may be utilized to improve the detection method's performance.

#### REFERENCES

- [1] Antoniou A (2019) Zao's deepfake face-swapping app shows uploading your photos is riskier than ever. The Conversation.
- [2] Perov I, Gao D, Chervoniy N, Liu K, Marangonda S, Umé C, Dpfks M, Facenheim SC, RP L, Jiang J, Zhang S, Wu P, Zhou B, Zhang W (2020) DeepFaceLab: a simple, flexible and extensible face swapping framework.
- [3] Masood M, Nawaz M, Malik KM, Javed A, Irtaza A (2021) Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward. arXiv preprint arXiv:2103.00484
- [4] Khalid H, Tariq S, Kim M, Woo SS (2021a) FakeAVCeleb: a novel audio-video multimodal deepfake dataset. arXiv preprint arXiv:2108.05080.
- [5] Jia Y, Zhang Y, Weiss RJ, Wang Q, Shen J, Ren F, Chen Z, Nguyen P, Pang R, Moreno IL, Wu Y (2018) Transfer learning from speaker verification to multispeaker text-to-speech synthesis. arXiv preprint arXiv:1806.04558.
- [6] Prajwal KR, Mukhopadhyay R, Namboodiri VP, Jawahar CV (2020, October) A lip sync expert is all you need for speech to lip generation in the wild. In Proceedings of the 28th ACM International Conference on Multimedia, pp 484-492.
- [7] Rossler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Nießner M (2019) Faceforensics++: learning to detect manipulated facial images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 1-11.
- [8] Kumar A, Bhavsar A, Verma R (2020) Detecting deepfakes with metric learning. In 2020 8th international workshop on biometrics and forensics (IWBF), IEEE, pp 1-6.
- [9] Singh P, Saha G, Sahidullah M (2021, January) Non-linear frequency warping using constant-Q transformation for speech emotion recognition. In 2021 International Conference on Computer Communication and Informatics (ICCCI), IEEE, pp 1-6.
- [10] Schörkhuber C, Klapuri A (2010, July) Constant-Q transform toolbox for music processing. In 7th sound and music computing conference, Barcelona, Spain, pp 3-64.
- [11] Delgado H, Todisco M, Sahidullah M, Sarkar AK, Evans N, Kinnunen T, Tan ZH (2016, December) Further optimisations of constant Q cepstral processing for integrated utterance and text-dependent speaker verification. In 2016 IEEE Spoken Language Technology Workshop (SLT), IEEE, pp 179-185.
- [12] Lidy T, Schindler A (2016, September) CQT-based convolutional neural networks for audio scene classification. In Proceedings of the detection and classification of acoustic scenes and events 2016 workshop (DCASE2016), IEEE Budapest, Vol 90, pp 1032-1048.
- [13] Halpern BM, Kelly F, van Son R, Alexander A (2020). Residual networks for resisting noise: analysis of an embeddings-based spoofing countermeasure.
- [14] Reimao RAM (2019) Synthetic speech detection using deep neural networks.
- [15] Li X, Li N, Weng C, Liu X, Su D, Yu D, Meng H (2021, June) Replay and synthetic speech detection with res2net architecture. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp 6354-6358.
- [16] Shen G, Tan Q, Zhang H, Zeng P, Xu J (2018) Deep learning with gated recurrent unit networks for financial sequence predictions. Procedia computer science 131: 895-903.
- [17] Zhou P, Shi W, Tian J, Qi Z, Li B, Hao H, Xu B (2016, August) Attention-based bidirectional long short-term memory networks for relation classification. In Proceedings of the 54th annual meeting of the association for computational linguistics, Vol 2, pp 207-212.
- [18] Cheng Q, Fu Y, Huang J et al (2022) Event detection based on the label attention mechanism. Int. J. Mach. Learn. & Cyber. <https://doi.org/10.1007/s13042-022-01655-y>.
- [19] Li Y, Lyu S (2018) Exposing deepfake videos by detecting face warping artifacts arXiv preprint arXiv:1811.00656.
- [20] Hu Z, Xie H, Wang Y, Li J, Wang Z, Zhang Y (2021, January) Dynamic Inconsistency-aware DeepFake Video Detection. In IJCAI.
- [21] Rana MS, Sung AH (2020, August) Deepfakestack: a deep ensemble-based learning technique for deepfake detection. In 2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom), IEEE, pp 70-75.
- [22] Ismail A, Elpeltagy M, S Zaki M, Eldahshan K (2021) A New deep learning-based methodology for video deepfake detection using XGBoost. Sensors 21(16): 5413.
- [23] Güera D, Delp EJ (2018, November) Deepfake video detection using recurrent neural networks. In 2018 15th IEEE international conference on advanced video and signal-based surveillance (AVSS), IEEE, pp 1-6.
- [24] Li Y, Chang MC, Lyu S (2018, December) In ictu oculi: exposing ai created fake videos by detecting eye blinking. In 2018 IEEE International Workshop on Information Forensics and Security (WIFS), IEEE, pp 1-7.
- [25] Zheng Y, Bao J, Chen D, Zeng M, Wen F (2021) Exploring temporal coherence for more general video face forgery detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 15044-15054.
- [26] Gu Z, Chen Y, Yao T, Ding S, Li J, Huang F, Ma L (2021, October) Spatiotemporal inconsistency learning for deepFake video detection. In Proceedings of the 29th ACM International Conference on Multimedia, pp 3473-3481.
- [27] Ismail A, Elpeltagy M, Zaki M, ElDahshan KA (2021) Deepfake video detection: YOLO-Face convolution recurrent approach. PeerJ Computer Science 7: e730.
- [28] Yang X, Li Y, Lyu S (2019, May) Exposing deep fakes using inconsistent head poses. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp 8261-8265.
- [29] Khalil SS, Youssef SM, Saleh SN (2021) iCaps-Dfake: an integrated capsule-based model for deepfake image and video detection. Future Internet 13(4): 93.
- [30] Ismail A, Elpeltagy M, Zaki MS, Eldahshan K (2022) An integrated spatiotemporal-based methodology for deepfake detection. Neural Comput & Applic. <https://doi.org/10.1007/s00521-022-07633-3>.
- [31] Witkowski M, Kacprzak S, Zelasko P, Kowalczyk K, Galka J (2017, August) Audio replay attack detection using high-frequency features. In Interspeech, pp 27-31.
- [32] Saranya MS, Padmanabhan R, Murthy HA (2018, July) Replay attack detection in speaker verification using non-voiced segments and decision level feature switching. In 2018 International Conference on Signal Processing and Communications (SPCOM), IEEE, pp 332-336.
- [33] Chen T, Kumar A, Nagarsheth P, Sivaraman G, Khoury E (2020, November) Generalization of audio deepfake detection. In Proc. Odyssey 2020 The Speaker and Language Recognition Workshop, pp 132-137.

- [34] Wu Z, Das RK, Yang J, Li H (2020) Light convolutional neural network with feature genuinization for detection of synthetic speech attacks. arXiv preprint arXiv:2009.09637.
- [35] Bartusiak ER, Delp EJ (2021) Frequency domain-based detection of generated audio. *Electronic Imaging* 2021(4): 273-1-273-7.
- [36] Zhang Y, Jiang F, Duan Z (2021) One-class learning towards synthetic voice spoofing detection. *IEEE Signal Processing Letters* 28: 937-941.
- [37] Gomez-Alanis A, Peinado AM, Gonzalez JA, Gomez AM (2019, September) A light convolutional GRU-RNN deep feature extractor for ASV spoofing detection. In *Proc. Interspeech*, Vol 2019, pp 1068-1072.
- [38] Huang L, Pun CM (2020) Audio replay spoof attack detection by joint segment-based linear filter bank feature extraction and attention-enhanced DenseNet-BiLSTM network. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28: 1813-1825.
- [39] Hua G, Bengiinteoh A, Zhang H (2021) Towards end-to-end synthetic speech detection. *IEEE Signal Processing Letters*.
- [40] Mittal T, Bhattacharya U, Chandra R, Bera A, Manocha D (2020, October) Emotions don't lie: an audio-visual deepfake detection method using affective cues. In *Proceedings of the 28th ACM international conference on multimedia*, pp 2823-2832.
- [41] Chugh K, Gupta P, Dhall A, Subramanian R (2020, October) Not made for each other: audio-visual dissonance-based deepfake detection and localization. In *Proceedings of the 28th ACM International Conference on Multimedia*, pp 439-447.
- [42] Gu Y, Zhao X, Gong C, Yi X (2020, November) Deepfake video detection using audio-visual consistency. In *International Workshop on Digital Watermarking*, Springer, Cham, pp 168-180.
- [43] Zhou Y, Lim SN (2021) Joint audio-visual deepfake detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp 14800-14809.
- [44] Gino J (2021) Audio-video deepfake detection through emotion recognition.
- [45] Mozley K (2021) Don't believe everything you hear: combining audio and visual cues for deepfake detection.
- [46] Blankertz B (2001) The constant Q transform. URL [http://doc. ml. tu-berlin. de/bbci/material/publications/Bla\\_constQ. Pdf](http://doc.ml.tu-berlin.de/bbci/material/publications/Bla_constQ.Pdf).
- [47] Schörkhuber C, Klapuri A, Sontacchi A (2012, September) Pitch shifting of audio signals using the constant-q transform. In *Proc. of the DAFx Conference*.
- [48] Todisco M, Delgado H, Evans NW (2016, June) A new feature for automatic speaker verification anti-spoofing: constant Q cepstral coefficients. In *Odyssey*, Vol 2016, pp 283-290.
- [49] Wang M, Wang R, Zhang XL, Rahardja S (2019, November) Hybrid constant-Q transform based CNN ensemble for acoustic scene classification. In *2019 Asia-pacific signal and information processing association annual summit and conference (APSIPA ASC)*, pp 1511-1516.
- [50] Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1251-1258.
- [51] Ghosh S, Pal A, Jaiswal S, Santosh KC, Das N, Nasipuri M (2019) SegFast-V2: Semantic image segmentation with less parameters in deep learning for autonomous driving. *International Journal of Machine Learning and Cybernetics* 10(11): 3145-3154.
- [52] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017, February) Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*.
- [53] Arfianti UI, Novitasari DCR, Widodo N, Hafiyusholeh M, Utami WD (2021) Sunspot number prediction using gated recurrent unit (GRU) algorithm. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)* 15(2): 141-152.
- [54] Le XH, Ho HV, Lee G (2019, September) Application of gated recurrent unit (GRU) network for forecasting river water levels affected by tides. In *International Conference on Asian and Pacific Coasts*, Springer, Singapore, pp 673-680.
- [55] Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E (2016, June) Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pp 1480-1489.
- [56] Hossin M, Sulaiman MN (2015) A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process* 5(2): 1.
- [57] Achlioptas P (2019) Stochastic gradient descent in theory and practice.
- [58] Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- [59] Ho Y, Wookey S (2019) The real-world-weight cross-entropy loss function: modeling the costs of mislabeling. *IEEE Access* 8: 4806-4813.
- [60] Khalid H, Kim M, Tariq S, Woo SS (2021b, October) Evaluation of an audio-video multimodal deepfake dataset using unimodal and multimodal detectors. In *Proceedings of the 1st workshop on synthetic multimedia-audiovisual deepfake generation and detection*, pp 7-15.
- [61] Aly M, Alotaibi NS (2022) A New Model to Detect COVID-19 Coughing and Breathing Sound Symptoms Classification from CQT and Mel Spectrogram Image Representation using Deep Learning. *International Journal of Advanced Computer Science and Applications*, pp 601-611.



# Three on Three Optimizer: A New Metaheuristic with Three Guided Searches and Three Random Searches

Purba Daru Kusuma, Ashri Dinimaharawati  
Computer Engineering, Telkom University, Bandung, Indonesia

**Abstract**—This paper presents a new swarm intelligence-based metaheuristic called a three-on-three optimizer (TOTO). This name is chosen based on its novel mechanism in adopting multiple searches into a single metaheuristic. These multiple searches consist of three guided searches and three random searches. These three guided searches are searching toward the global best solution, searching for the global best solution to avoid the corresponding agent, and searching based on the interaction between the corresponding agent and a randomly selected agent. The three random searches are the local search of the corresponding agent, the local search of the global best solution, and the global search within the entire search space. TOTO is challenged to solve the classic 23 functions as a theoretical optimization problem and the portfolio optimization problem as a real-world optimization problem. There are 13 bank stocks from Kompas 100 index that should be optimized. The result indicates that TOTO performs well in solving the classic 23 functions. TOTO can find the global optimal solution of eleven functions. TOTO is superior to five new metaheuristics in solving 17 functions. These metaheuristics are grey wolf optimizer (GWO), marine predator algorithm (MPA), mixed leader-based optimizer (MLBO), golden search optimizer (GSO), and guided pelican algorithm (GPA). TOTO is better than GWO, MPA, MLBO, GSO, and GPA in solving 22, 21, 21, 19, and 15 functions, respectively. It means TOTO is powerful to solve high-dimension unimodal, multimodal, and fixed-dimension multimodal problems. TOTO performs as the second-best metaheuristic in solving a portfolio optimization problem.

**Keywords**—*Optimization; metaheuristic; swarm intelligence; portfolio optimization; Kompas 100; bank*

## I. INTRODUCTION

Many real-world problems can be seen as optimization problems. This circumstance comes from the nature of human behavior or activity in achieving their objective most efficiently. Ironically, people always find certain limitations or constraints. This consideration is the same as the optimization work. In general, optimization is constructed by two elements: objective and constraint. In the optimization problem, many solutions can be chosen in the solution space. However, some solutions are better than others. One solution that is the best one is called the optimal global solution.

The objective of optimization can be minimization or maximization. In the minimization, the optimal global solution is the solution with the lowest value. Some experimental parameters in the minimization, such as delay [1], total order completion time [2], idle time [2], tardiness cost and maintenance [3], project duration [4], energy consumption [5], transmission losses [6], and so on. On the other hand, in

maximization, the optimal global solution is the solution with the highest value. Some parameters in the maximization are profit [7], voltage stability [6], revenue [8], service level [9], and so on.

Two ways can be chosen to solve the optimization problem. These methods are a mathematical method and a metaheuristic [10]. The mathematical method is robust in solving a simple optimization problem. It guarantees finding the optimal global solution. However, the mathematical or deterministic method often fails to solve a complex optimization problem, such as a non-convex or multimodal problem. Moreover, the mathematical method is less flexible in facing various real-world optimization problems [11].

On the other hand, metaheuristic is widely used in many optimizations. Metaheuristics have several advantages. First, it is flexible enough to be implemented in various problems because it focuses on the objectives and constraints [10]. Second, it can be implemented in an environment with limited computational resources because of its approximate approach so that not all possible solutions are traced [12]. This approximate approach comes with the consequence that metaheuristic does not guarantee finding the optimal global solution.

There are hundreds of metaheuristics developed in recent decades. Moreover, many metaheuristics have been developed in recent years. Many of them used metaphors for their name. Many of these shortcoming metaheuristics were inspired by animal behavior, such as the butterfly optimization algorithm (BOA) [13], chameleon swam algorithm (CSA) [14], coati optimization algorithm (COA) [15], Komodo mlipir algorithm (KMA) [16], northern goshawk optimizer (NGO) [17], raccoon optimization algorithm (ROA) [18], marine predator algorithm (MPA) [19], Tasmanian devil optimizer (TDO) [11], snake optimizer (SO) [10], white shark optimizer (WSO) [20], guided pelican algorithm (GPA) [21], and so on. Some metaheuristics were inspired by the mechanics of plants, such as the tunicate swarm algorithm (TSA) [22], flower pollination algorithm (FPA) [23], and so on. Some metaheuristics are named based on their references in their guided search, such as three influential member-based optimizers (TIMBO) [24], mixed leader-based optimizers (MLBO) [25], multi-leader optimizers (MLO) [26], random selected leader-based optimizer (RLSBO) [27], and so on. Some metaheuristics were inspired by human activities, such as stochastic paint optimizer (SPO) [28], modified social forces algorithm (MSFA) [29], driving training-based optimizer (DTBO) [30], and so on. Some metaheuristics were free from metaphor and named based on

their central concept, such as average subtraction-based optimizer (ASBO) [31], golden search optimizer (GSO) [32], total interaction algorithm (TIA) [33], and so on.

Many early metaheuristics depend on a single strategy to find the optimal solution to problems. Unfortunately, as mentioned in the no-free-lunch theory, no metaheuristics can solve all optimization problems [34]. A search may be excellent for solving some optimization problems [34]. Meanwhile, the performance of this search may need to improve in solving other optimization problems. This circumstance gives strong motivation to develop multi search-based metaheuristic. Many shortcomings in metaheuristics were built by accommodating multiple searches in every iteration. Moreover, this multiple search strategy is proven better in solving an optimization problem. However, adding more searches in a single metaheuristic can increase the algorithm complexity and, in the end, the computational resources.

Many existing swarm-based metaheuristics depend more on the guided search than the random one. In some metaheuristics like KMA [16] or COA [15], segregation of roles is implemented in the population. It means that some agents implement segregation of roles while some other agents do not implement segregation of roles. A more guided search is deployed in many swarm-based metaheuristics with multiple searches. Some metaheuristic deploys neighborhood search only or does not implement any random search. Meanwhile, random search is essential in tackling the local optimal entrapment often found in multimodal problems.

Based on this problem and motivated by the no-free-lunch theory, this paper is aimed to propose a new swarm-based metaheuristic with multiple searches where there is a balance between the guided search and the random search. This proposed metaheuristic is named a three-on-three optimizer (TOTO). This name represents the six searches adopted in this metaheuristic. Three searches are guided searches, while three other searches are random.

The scientific contributions of this paper are listed as follows.

- 1) This paper proposes a new swarm-based metaheuristic with multiple searches within this metaheuristic.
- 2) This paper proposes a balanced proportion between the guided search and random search in every iteration.
- 3) The evaluation is performed by benchmarking the performance of the proposed TOTO with five shortcoming metaheuristics: GWO, MPA, MLBO, GSO, and GPA.
- 4) The benchmark test to evaluate the proposed metaheuristic in solving the classic 23 functions and the comparison with other shortcoming metaheuristics is carried out as a proof of concept regarding the significant performance of TOTO.
- 5) The stock optimization problem is chosen as evaluation of the proposed TOTO in solving real-world optimization problem in financial sector and comparison with other metaheuristics.

The remainder of this paper is arranged as follows. The investigation of several shortcomings of metaheuristics is carried out in section two. This investigation includes the mechanics adopted in these metaheuristics and the position of this work to make the novelty and contribution of this work clear. A detailed presentation of the proposed metaheuristic can be found in section three. This presentation includes the concept, algorithm, mathematical model, and algorithm complexity. The test carried out in this work regarding the evaluation of the proposed metaheuristic is presented in section four. The investigation of the result and findings are explored in section five. Finally, the conclusion and the opportunity for future works are summarized in section six.

## II. RELATED WORKS

Metaheuristics have been evolving for decades. Early metaheuristics are simple with a single strategy. In general, they deploy neighborhood search with some strategy to avoid the local optimal. Simulated annealing (SA) and tabu search (TS) are examples of metaheuristics that adopt a simple neighborhood search with a distinct approach to tackling the optimal local problem. In simulated annealing, some probabilistic calculations can still accept a worse solution, where the acceptance becomes more difficult as the iteration goes [35]. Meanwhile, tabu search uses some list to avoid old solutions will be revisited for a certain period [36].

The evolution goes with the development of population-based metaheuristics. Genetic algorithm (GA) and invasive weed optimizer (IWO) are clear examples of these metaheuristics. Population-based metaheuristics give two main advantages. First, they improve the solution faster than the single solution-based metaheuristics. Second, they give broader tracing within the search space. GA has a simple solution in improving the solution by implementing a crossover strategy between two solutions and mutation for exploration [37]. In IWO, multiple new solutions are generated around every existing solution based on normal distribution [38]. Then, all solutions are ranked, and the worse ones are eliminated [38].

The evolution continues with the introduction of a swarm-based metaheuristic. A particle swarm optimizer (PSO) is an example of an early swarm-based metaheuristic. In the swarm-based metaheuristic, each solution can be seen as an autonomous agent that does not have centralized coordination. However, specific interaction and collective intelligence are conducted to help the improvement. As a member of the swarm, each agent conducts a guided search. It means each agent moves in a specific direction, called a guided search toward some references. As an early swarm-based metaheuristic, PSO deploys a simple strategy. In PSO, each agent moves toward the reference at a certain speed [39]. Its references are the combination between the local best solution and the global best solution [39].

In recent decades, the development of swarm-based metaheuristics is more extensive. Various aspects can be used as a baseline for developing new swarm-based metaheuristics. These aspects include reference, stochastic movement, acceptance rejection, segregation of roles, etc. Moreover, many shortcomings of metaheuristics deploy multiple strategies rather than a single strategy to improve its quality in achieving

a better final solution or achieving the objective faster. This multiple strategy/search approach is carried out in a single or multiple phases. Some metaheuristics are also equipped with random search to avoid the convergence achieved faster and

trapped in the local optimal. A detailed review of shortcoming swarm-based metaheuristics is presented in Table I. The mechanics of the proposed metaheuristic is presented in the last row to make a clear position and contribution of this work.

TABLE I. DETAIL SUMMARY OF SEVERAL SHORTCOMING METAHEURISTIC

No	Metaheuristic	Number of Phases	Number of Searches	Guided Searches	Random Searches	Segregation of Roles
1	COA [15]	2	3	searching toward the best solution in the current iteration; searching toward a randomly solution within the search space	searching around the local search space	yes
2	MSFA [29]	1	2	searching toward a target where the target is a random solution within the search space	searching around the local search	no
3	CSA [14]	3	4	searching toward the gap between the local best solution and global best solution; searching toward the mixture between the global best solution and local best solution	searching within the search space; searching within the local search space;	yes
4	KMA [16]	1	4	searching toward the resultant of several best solutions; searching toward the resultant of several better best solutions and avoiding the resultant of several worse best solutions; and searching toward the best solution in the current iteration.	searching within the search space	yes
5	SO [10]	2	4	searching of the best solution to avoid the corresponding solution; searching toward the best solution in different group; searching toward a solution outside the group with same index	searching within the search space	yes
6	ASBO [31]	3	3	searching relative to the average between the best and worst solutions in the current iteration; search toward the gap between the best and worst solutions in the current iteration; searching toward the gap between the corresponding solution and the best solution	-	no
7	TIA [33]	1	1	searching relative to all other solutions in the population	-	no
8	WSO [20]	2	3	searching toward the mixture between the global best solution and local best solution; searching of the best solution to avoid the corresponding solution.	searching between the previous solution and its new guided search.	yes
9	GSO [32]	1	1	searching toward the mixture between the local best solution and the global best solution	-	no
10	NGO [17]	2	2	searching toward a randomly selected solution	search within the local search space.	no
11	GPA [21]	1	2	searching toward the global best solution	searching within the local search space	no
12	MLBO [25]	1	1	searching toward the mixture between the best solution in the current iteration and a randomly selected solution	-	no
13	TIMBO [24]	3	3	searching toward the best solution in the current iteration; searching to avoid the worst solution in the current iteration; and searching relative to the mean member in the current iteration	-	no
14	SPO [28]	1	4	searching toward adjacent solutions; searching toward the gap between a solution from the primary group and a solution from the tertiary group; searching toward the average among a solution from each group; and searching toward the average of a solution from each group plus a randomly selected solution	-	no
15	RSLBO [27]	1	1	searching toward a randomly selected solution	-	no
16	this work	1	6	searching toward the global best solution; searching of the global best solution to avoid the corresponding solution; searching relative to a randomly selected solution;	searching within the local search space of the corresponding solution; searching within the local search space of the global best solution; searching within the search space	no

Table I indicates the limitations of the existing swarm-based metaheuristics. First, many swarm based metaheuristics prioritize guided search rather than random search. Second, some swarm-based metaheuristics deploys only one random search while some others do not deploy any random search. Third, searching methods implemented in every metaheuristic are still less than five methods. Fourth, these multiple searches are performed in the multiple phase process. Based on this circumstance, there is a room in developing swarm-based metaheuristic that deploys multiple searches and puts the balance between guided search and random search as in this work.

### III. PROPOSED MODEL

TOTO is a swarm-based metaheuristic with multiple strategies in every iteration. The multiple-strategy approach aims to cover each strategy's disadvantages because each has its strengths and weaknesses. The multiple strategy approach is interpreted by conducting six searches by each agent in every iteration. It means that these searches are mandatory for each agent. In other words, TOTO does not implement segregation of roles. It differs from other metaheuristics, such as RDA, KMA, or SKA, where the segregation of roles is implemented among the population. These searches are three guided searches and three random searches. Each search generates a candidate. Then, the best candidate among these six candidates is chosen as the selected candidate to be compared with the corresponding agent. If this candidate is better than the current solution of the corresponding agent, this candidate becomes the replacement as a new solution for the corresponding agent. This mechanism is also different from some metaheuristics, such as DTBO, GPA, NGO, and so on, where the searches are conducted sequentially. This new solution is then compared with the current global best solution. If this new solution is better than the global best solution, then the global best solution is updated.

The global best solution and a randomly selected agent become the reference in these three guided searches. In the first guided search, a candidate is generated based on the movement of the corresponding agent toward the best global solution. In

the second guided search, a candidate is generated based on the movement of the global best solution avoiding the corresponding solution. In the third guided search, a candidate is generated based on the movement of the corresponding agent relative to a randomly selected agent. If the randomly selected agent is better than the corresponding agent, then the third candidate is generated based on the movement of the corresponding agent toward the randomly selected agent. Otherwise, the third candidate is generated based on the movement of the corresponding agent, avoiding the randomly selected agent.

The local search space is used for the first and second random searches. On the other hand, local search space is not needed in the third random search. The local search space width declines linearly due to the increase in the iteration. In the first random search, a candidate is generated within the local search space of the corresponding agent. In the second search space, a candidate is generated within the local search space of the global best solution. A candidate is generated within the search space in the third random search. The illustration of these six searches is presented in Fig. 1.

Based on the previous explanation, the rationale of the proposed strategy is highlighted and summarized as follows.

- The multiple searches are proven better than single search as the multiple search approach is adopted in various shortcoming metaheuristics.
- The balance between the guided searches and random searches is designed to give balance between exploration capability and exploitation capability.
- Multiple references are adopted to expand the searching capability because searching process cannot depend on only single reference.
- The strict acceptance-rejection approach is adopted to avoid the searching process moves to the worse solution or area.

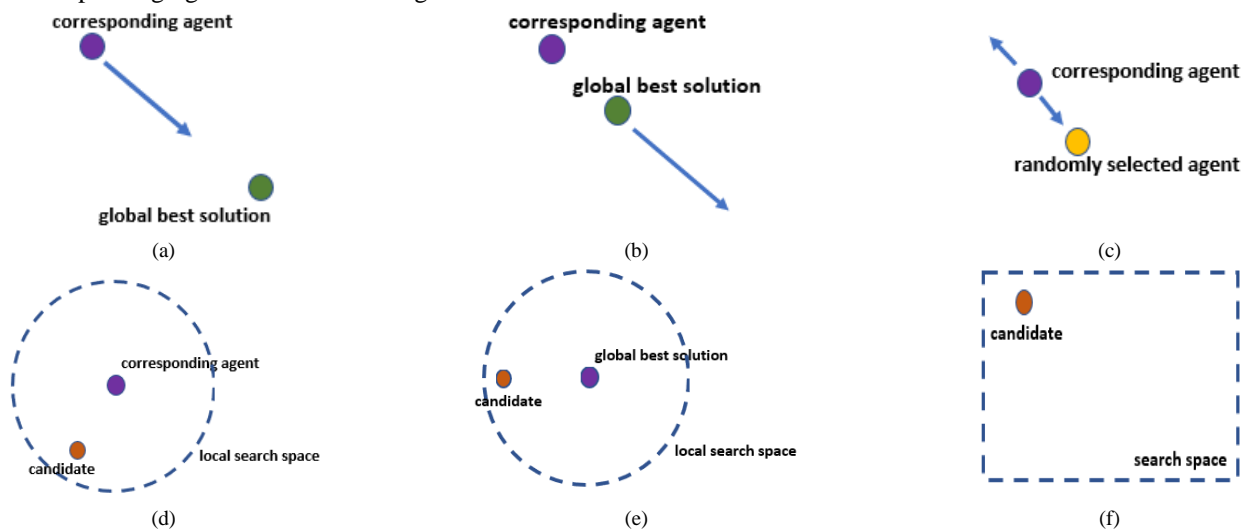


Fig. 1. Illustration of six searches in TOTO: (a) First guided search, (b) Second guided search, (c) Third guided search, (d) First random search, (e) Second random search, and (f) Third random search.

The algorithm of TOTO is then constructed based on its main concept as the formalization of this metaheuristic. This algorithm is presented in algorithm 1. Equation (1) to (11) is used as formalization of the related process. The annotations used in this paper are as follows.

$a$	agent
$A$	set of agents
$a_{best}$	global best agent
$a_{sel}$	randomly selected agent
$a_l$	lower boundary
$a_u$	upper boundary
$c_1$	first candidate
$c_2$	second candidate
$c_3$	third candidate
$c_4$	fourth candidate
$c_5$	fifth candidate
$c_6$	sixth candidate
$c_{sel}$	selected candidate
$f$	fitness function
$r_1$	real uniform random number between 0 and 1
$r_2$	integer uniform random number between 1 and 2
$r_3$	real uniform random number between -1 and 1
$t$	iteration
$t_{max}$	maximum iteration
$u$	uniform random

**algorithm 1:** main algorithm of TOTO

```

1  Begin
2  for all  $a$  in  $A$ 
3    generate  $a$  using (1)
4    update  $a_{best}$  using (2)
5  end for
6  for  $t = 1$  to  $t_{max}$ 
7    for all  $a$  in  $A$ 
8      first guided search using (3)
9      second guided search using (4)
10     third guided search using (5) and (6)
11     first random search using (7)
12     second random search using (8)
13     third random search using (9)
14     choose  $c_{sel}$  using (10)
15     update  $a$  using (11)
16     update  $a_{best}$  using (2)
17   end for
18 end for
19 end

```

$$a = U(a_l, a_u) \quad (1)$$

$$a_{best}' = \begin{cases} a, & f(a) < f(a_{best}) \\ a_{best}, & \text{else} \end{cases} \quad (2)$$

$$c_1 = a + r_1(a_{best} - r_2a) \quad (3)$$

$$c_2 = a_{best} + r_1(a_{best} - r_2a) \quad (4)$$

$$a_{sel} = U(A) \quad (5)$$

$$c_3 = \begin{cases} a + r_1(a_{sel} - r_2a), & f(a_{sel}) < f(a) \\ a + r_1(a - r_2a_{sel}), & \text{else} \end{cases} \quad (6)$$

$$c_4 = a + r_3 \left(1 - \frac{t}{t_m}\right) (a_u - a_l) \quad (7)$$

$$c_5 = a_{best} + r_3 \left(1 - \frac{t}{t_m}\right) (a_u - a_l) \quad (8)$$

$$c_6 = a_l + r_1(a_u - a_l) \quad (9)$$

$$c_{sel} = c \in C, \min(f(c)) \quad (10)$$

$$a' = \begin{cases} c_{sel}, & f(c_{sel}) < f(a) \\ a, & \text{else} \end{cases} \quad (11)$$

The explanation of (1) to (11) is as follows. Equation (1) states that the initial solution is randomized within the search space. Equation (2) states that the global best solution is updated by comparing the current value of the global best solution and the new value of the corresponding agent. If the corresponding agent is better than the global best solution, then the corresponding agent replaces the current global best solution. Equation (3) states that the candidate of the first guided search is generated based on the movement of the corresponding agent toward the global best solution. Equation (4) states that the candidate of the second guided search is generated based on the movement of the global best solution avoiding the corresponding agent. Equation (5) states that an agent is randomly selected among the population. Equation (6) states that the candidate of the third guided search is generated based on the movement of the corresponding agent relative to a randomly selected agent. Equation (7) states that the candidate of the first random search is generated based on the neighborhood search around the corresponding agent. Equation (8) states that the candidate of the second random search is generated based on the neighborhood search around the global best solution. Equation (9) states that the candidate of the third random search is generated based on the constant random search within the search space. Equation (10) states that the best candidate among these six candidates is selected to be compared with the corresponding solution. Equation (11) states that this candidate replaces the corresponding solution if this candidate is better than the corresponding solution.

The algorithm complexity of TOTO is presented as  $O(6t_{max}.n(A))$ . Based on this presentation, the complexity of TOTO is affected by two parameters: maximum iteration and population size. Each parameter has a linear proportion to the complexity.

#### IV. SIMULATION AND RESULT

This section discusses the performance analysis regarding the proposed metaheuristic. Some optimization tests are carried out due to provide the performance data. The first test is carried out to evaluate the performance of TOTO in solving classic 23 functions. These functions are standard as benchmark tests in many works proposing new metaheuristics. These functions consist of seven high dimension unimodal functions (Sphere, Schwefel 2.22, Schwefel 1.2, Schwefel 2.21, Rosenbrock, Step, and Quartic), six high-dimension multimodal functions (Schwefel, Rastrigin, Ackley, Griewank, Penalized, and Penalized 2), and ten fixed dimension multimodal functions (Shekel Foxholes, Kowalik, Six Hump Camel, Branin, Goldstein-Price, Hartman 3, Hartman 6, Shekel 5, Shekel 7, and Shekel 10). The second test is carried out to analyze the

hyperparameters of TOTO. The third test is carried out to evaluate the performance of TOTO in solving the practical optimization problem. The third test is carried out to evaluate the performance of TOTO in solving a real-world optimization problem. The portfolio optimization problem is chosen as the use case in this work.

The first test compares TOTO with five shortcoming swarm-based metaheuristics: GWO, MPA, MLBO, GSO, and GPA. GWO and MPA are older than the others but widely used in various optimization works. On the other hand, MLBO, GSO, and GPA are newer, but the use of these metaheuristics is still rare.

Several parameters regarding the first test are set as follows. The population size is 10. The maximum iteration is 50. The problem dimension is 50. The fishing aggregate devices for MPA are 0.5. The number of candidates for GPA is 5. The result is presented in Table II. The average fitness score, less than  $10^{-4}$ , is rounded to the nearest  $10^{-4}$  value. The best score in every function is presented in bold font. The cluster-based comparison is presented in Table III.

Table II indicates the excellent performance of TOTO in solving the 23 classic functions and competing with five other metaheuristics. The result presents that TOTO can find the

optimal global solution of eight functions (Sphere, Schwefel 2.22, Schwefel 2.21, Rastrigin, Ackley, Six Hump Camel, and Goldstein-Price). Moreover, TOTO also performs as the best metaheuristics in solving 17 functions. Seven high-dimension unimodal functions (Sphere, Schwefel 2.22, Schwefel 1.2, Schwefel 2.21, Rosenbrock, Step, and Quartic), five high-dimension multimodal functions (Rastrigin, Ackley, Griewank, Penalized, and Penalized 2), and five fixed dimension multimodal functions (Kowalik, Six Hump Camel, Branin, Goldstein-Price, and Shekel 7). But other metaheuristics also find the optimal global solutions in solving three functions (Schwefel 2.22, Six Hump Camel, and Branin).

Table III strengthens the superiority of TOTO. TOTO is better than GWO, MPA, MLBO, GSO, and GPA in solving 22, 21, 21, 19, and 15 functions, respectively. GWO becomes the easiest metaheuristic to beat. On the other hand, GPA becomes the most challenging metaheuristic to beat. Compared to GPA, TOTO is superior in solving high-dimension functions. On the other hand, GPA is superior in solving fixed-dimension functions.

In the second test, the performance of TOTO is evaluated due to the increase in the maximum iteration. This test has three maximum iterations: 100, 150, and 200. The result is presented in Table IV.

TABLE II. SIMULATION RESULT (AVERAGE FITNESS SCORE)

Function	GWO	MPA	MLBO	GSO	GPA	TOTO	Better than
Sphere	7.4328x10 <sup>2</sup>	2.8092x10 <sup>3</sup>	1.5375x10 <sup>4</sup>	2.2247x10 <sup>4</sup>	1.5742x10 <sup>2</sup>	<b>0.0000</b>	GWO, MPA, MLBO, GSO, GPA
Schwefel 2.22	0.0000	0.0000	0.0000	9.9113x10 <sup>63</sup>	0.0000	<b>0.0000</b>	GSO
Schwefel 1.2	2.6197x10 <sup>4</sup>	1.9080x10 <sup>4</sup>	3.8128x10 <sup>4</sup>	4.5054x10 <sup>4</sup>	8.5787x10 <sup>3</sup>	<b>0.0003</b>	GWO, MPA, MLBO, GSO, GPA
Schwefel 2.21	1.6220x10 <sup>1</sup>	2.2948x10 <sup>1</sup>	4.1616x10 <sup>1</sup>	4.0027x10 <sup>1</sup>	2.2449x10 <sup>1</sup>	<b>0.0000</b>	GWO, MPA, MLBO, GSO, GPA
Rosenbrock	2.2470x10 <sup>6</sup>	9.1381x10 <sup>5</sup>	1.0097x10 <sup>7</sup>	1.9217x10 <sup>7</sup>	7.9361x10 <sup>3</sup>	<b>4.8833x10<sup>1</sup></b>	GWO, MPA, MLBO, GSO, GPA
Step	9.2781x10 <sup>2</sup>	3.0696x10 <sup>3</sup>	1.5544x10 <sup>4</sup>	2.0205x10 <sup>4</sup>	1.4096x10 <sup>2</sup>	<b>9.0134</b>	GWO, MPA, MLBO, GSO, GPA
Quartic	0.8404	0.7544	8.0704	1.6288x10 <sup>1</sup>	0.3552	<b>0.0037</b>	GWO, MPA, MLBO, GSO, GPA
Schwefel	-1.2504x10 <sup>1</sup>	-3.6788x10 <sup>3</sup>	-5.125x10 <sup>3</sup>	-5.1120x10 <sup>3</sup>	<b>-1.1075x10<sup>4</sup></b>	-5.648x10 <sup>3</sup>	GWO, MPA, MLBO, GSO
Rastrigin	1.3626x10 <sup>1</sup>	3.1846x10 <sup>2</sup>	3.8775x10 <sup>2</sup>	3.7542x10 <sup>2</sup>	1.8105x10 <sup>2</sup>	<b>0.0000</b>	GWO, MPA, MLBO, GSO, GPA
Ackley	2.1327	9.1212	1.5494x10 <sup>1</sup>	1.9245x10 <sup>1</sup>	4.5612	<b>0.0000</b>	GWO, MPA, MLBO, GSO, GPA
Griewank	1.6133	2.4440x10 <sup>1</sup>	1.4181x10 <sup>2</sup>	2.0615x10 <sup>2</sup>	2.2037	<b>0.0021</b>	GWO, MPA, MLBO, GSO, GPA
Penalized	2.0545x10 <sup>5</sup>	9.0031x10 <sup>2</sup>	1.5497x10 <sup>6</sup>	1.1357x10 <sup>7</sup>	1.6725x10 <sup>1</sup>	<b>0.6670</b>	GWO, MPA, MLBO, GSO, GPA
Penalized 2	3.0459x10 <sup>7</sup>	3.4758x10 <sup>5</sup>	1.9640x10 <sup>7</sup>	4.4838x10 <sup>7</sup>	4.4657x10 <sup>1</sup>	<b>2.9620</b>	GWO, MPA, MLBO, GSO, GPA
Shekel Foxholes	2.3602x10 <sup>1</sup>	8.1399	3.9583	9.3023	<b>1.0927</b>	1.1388	GWO, MPA, MLBO, GSO
Kowalik	0.1970	0.0121	0.0046	0.0150	0.0053	<b>0.0015</b>	GWO, MPA, MLBO, GSO, GPA
Six Hump Camel	0.2901	-1.0037	-1.0235	-1.0316	-1.0316	<b>-1.0316</b>	GWO, MPA, MLBO
Branin	5.5155x10 <sup>1</sup>	0.9437	0.4016	0.3981	0.3981	<b>0.3981</b>	GWO, MPA, MLBO
Goldstein-Price	6.0122x10 <sup>2</sup>	8.5646	3.8505	4.2277	3.0001	<b>3.0000</b>	GWO, MPA, MLBO, GSO, GPA
Hartman 3	-0.0010	<b>-3.5541</b>	-0.0495	-0.0426	-0.0495	-0.0495	GWO, GSO
Hartman 6	-0.0051	-1.5787	-3.1239	-2.9718	<b>-3.2998</b>	-3.2103	GWO, MPA, MLBO, GSO
Shekel 5	-0.2731	-1.0622	-3.5332	<b>-6.6056</b>	-5.6785	-6.5205	GWO, MPA, MLBO, GPA
Shekel 7	-0.2936	-1.3635	-5.5027	-5.3298	-8.1483	<b>-7.6367</b>	GWO, MPA, MLBO, GSO, GPA
Shekel 10	-0.3217	-1.4165	-4.3085	-6.0164	<b>-9.5547</b>	-5.7171	GWO, MPA, MLBO

TABLE III. CLUSTER BASED COMPARISON

Cluster	GWO	MPA	MLBO	GSO	GPA
1	6	6	6	7	6
2	6	6	6	6	5
3	10	9	9	6	4
Total	22	21	21	19	15

TABLE IV. RELATION BETWEEN MAXIMUM ITERATION AND AVERAGE FITNESS SCORE

Function	Average Fitness Score		
	$t_{max} = 100$	$t_{max} = 150$	$t_{max} = 200$
Sphere	0.0000	0.0000	0.0000
Schwefel 2.22	0.0000	0.0000	0.0000
Schwefel 1.2	0.0000	0.0000	0.0000
Schwefel 2.21	0.0000	0.0000	0.0000
Rosenbrock	$4.8843 \times 10^1$	$4.8848 \times 10^1$	$4.8845 \times 10^1$
Step	8.7200	8.5779	8.9453
Quartic	0.0011	0.0008	0.0003
Schwefel	$-6.4723 \times 10^3$	$-6.7829 \times 10^3$	$-7.3773 \times 10^3$
Rastrigin	0.0000	0.0000	0.0000
Ackley	0.0000	0.0000	0.0000
Griewank	0.0000	0.0000	0.0000
Penalized	0.6270	0.5446	0.5199
Penalized 2	2.9369	2.8683	2.8744
Shekel Foxholes	1.1869	0.9980	0.9980
Kowalik	0.0015	0.0015	0.0006
Six Hump Camel	-1.0316	-1.0316	-1.0316
Branin	0.3981	0.3981	0.3981
Goldstein-Price	3.0000	3.0000	3.0000
Hartman 3	-0.0495	-0.0495	-0.0495
Hartman 6	-3.2681	-3.2221	-3.2169
Shekel 5	-7.3673	-7.7048	-8.2857
Shekel 7	-8.1151	-8.8114	-8.8197
Shekel 10	-8.5191	-8.4954	-9.0716

Table IV indicates that the increase in maximum iteration does not improve the performance of TOTO in most functions. There are two possible reasons for this circumstance. The first reason is the optimal global solution has been found when the maximum iteration is still low. The second reason is TOTO fails to improve, although the optimal global solution has yet to be found. In some functions, the increase of maximum iteration improves the performance of TOTO but is less significant. Fortunately, when the maximum iteration increases, TOTO can find the optimal global solution of Schwefel 1.2, Griewank, and Shekel Foxholes.

In the third test, the performance of TOTO is evaluated due to the increase of population size. There are three values of population size: 20, 30, and 40. The result is presented in Table V.

TABLE V. RELATION BETWEEN POPULATION SIZE AND AVERAGE FITNESS SCORE

Function	Average Fitness Score		
	$n(S) = 20$	$n(S) = 30$	$n(S) = 40$
Sphere	0.0000	0.0000	0.0000
Schwefel 2.22	0.0000	0.0000	0.0000
Schwefel 1.2	0.0003	0.0000	0.0000
Schwefel 2.21	0.0000	0.0000	0.0000
Rosenbrock	$4.8806 \times 10^1$	$4.8727 \times 10^1$	$4.8723 \times 10^1$
Step	8.0915	7.7506	7.2435
Quartic	0.0027	0.0012	0.0011
Schwefel	$-6.0975 \times 10^3$	$-6.746 \times 10^1$	$-6.8492 \times 10^3$
Rastrigin	0.0000	0.0000	0.0000
Ackley	0.0000	0.0000	0.0000
Griewank	0.0006	0.0000	0.0000
Penalized	0.5811	0.5286	0.4791
Penalized 2	2.8680	2.7882	2.8189
Shekel Foxholes	0.9980	0.9980	0.9980
Kowalik	0.0015	0.0006	0.0009
Six Hump Camel	-1.0316	-1.0316	-1.0316
Branin	0.3981	0.3981	0.3981
Goldstein-Price	3.0000	3.0000	3.0000
Hartman 3	-0.0495	-0.0495	-0.0495
Hartman 6	-3.2551	-3.2150	-3.2796
Shekel 5	-7.8440	-8.1561	-8.0214
Shekel 7	-8.1235	-8.7475	-9.3842
Shekel 10	-7.4649	-9.1384	-9.4117

Table V indicates there are also two circumstances regarding the increase in the population size. First, the performance of TOTO needs to improve the increase in population size. Like in the second test, this circumstance happens because the optimal global solution has been found or TOTO fails to improve, although the optimal global solution has yet to be found. Like in the first test, the population size increase also makes TOTO find the optimal global solution of three more functions: Schwefel 1.2, Griewank, and Shekel Foxholes. Meanwhile, in some other functions, TOTO can improve its performance.

In the fourth test, TOTO is challenged to solve the portfolio optimization problem as a real-world problem. In this work, the optimization determines the stocks the investor should buy. The stocks selected in this work are stocks from the banking sector in Indonesia, which are listed in Kompas 100 index. Kompas 100 is a list that consists of preferred stocks in Indonesia. 13 stocks in the banking sector are listed in Kompas 100. Detailed information regarding these stocks is presented in Table VI. In Table VI, the second column represents the stock index, the third column represents the stock price taken on 11 November 2022, and the third column represents the six-month capital gain of these stocks. The price and capital gain are presented in rupiah per share.

TABLE VI. STOCK INFORMATION

No	Stock Index	Price	Capital Gain
1	AGRO	525	-410
2	AMAR	366	42
3	ARTO	5,700	-2,100
4	BACA	136	-23
5	BBCA	8,725	1,350
6	BBHI	2,380	-2,450
7	BBNI	9,175	325
8	BBRI	4,590	260
9	BTPN	2,620	90
10	BMRI	10,100	2,225
11	BNBA	1,530	-310
12	BRIS	1,360	-55
13	BTPS	3,060	110

The objective of this portfolio optimization is to maximize the total capital gain. The total capital gain is calculated by accumulating the capital gain from the stocks. Two constraints limit this optimization. The first constraint is that the investor can purchase stock from 100 to 1,000 lots for each stock. One lot means 100 shares. The second constraint is that the maximum investment is 4,000,000,000 rupiah. This work benchmarked TOTO with GPA, GSO, MLBO, MPA, and GWO. The population size is ten, and the maximum iteration is 50. The result is presented in Table VII.

TABLE VII. PORTFOLIO OPTIMIZATION RESULT

No	Metaheuristic	Total Capital Gain
1	TOTO	382,157,681
2	GPA	384,955,295
3	GSO	289,744,968
4	MLBO	361,041,190
5	MPA	188,953,890
6	GWO	-16,222,768

Table VII indicates the good performance of TOTO. It is shown that TOTO outperforms GSO, MLBO, MPA, and GWO. Meanwhile, GPA is better than TOTO with very narrow gap. GWO becomes the metaheuristic with the poorest performance because it produces negative total capital gain which means loss.

## V. DISCUSSION

This fifth section will discuss the more profound analysis of the result and findings. This discussion consists of four parts. The first part is the discussion regarding the performance of TOTO in solving the 23 classic functions. The second part is the discussion regarding hyperparameters. The third part is the discussion regarding the performance of TOTO in solving the portfolio optimization problem. The fourth part is the discussion regarding the limitation of this work, especially the proposed metaheuristic.

The first discussion is about the performance of TOTO. TOTO performs well in solving the classic 23 functions and benchmarking with five other metaheuristics. As mentioned previously, TOTO performs as the best metaheuristic in solving 17 functions in the low population size and low maximum iteration circumstances. Meanwhile, TOTO can find the optimal global solution of eight functions in these circumstances. Moreover, TOTO can find the optimal global solution of three additional functions in the high maximum iteration or population size circumstances. Comparing metaheuristics, TOTO is better than GWO, MPA, MLBO, and GSO in almost all functions. Meanwhile, TOTO is better than GPA in 15 functions.

In general, the superiority of TOTO occurs in all three groups of these 23 functions. It means that TOTO can tackle various problems indicated in these functions. Due to the consideration that multimodal functions are used to test the exploration capability while unimodal functions are used to test the exploitation capability [40], the proposed TOTO is proven in having superior exploration and exploitation capabilities. Meanwhile, the inferiority of TOTO in solving the fixed dimension multimodal functions compared to GPA is that GPA performs very well in solving the fixed dimension multimodal problems. But the superiority of GPA in solving these functions comes with the consequence that the complexity of GPA is higher than TOTO. GPA generates several candidates in its searches, whether it's guided search toward the global best solution or its local search [21]. On the other hand, the tournament of six searches adopted in TOTO is proven better than the multiple candidate strategy adopted in GPA [21]. The result also proves that the sorting mechanism at the beginning of the iteration, as adopted in GWO [41], is unimportant.

The second discussion is related to hyperparameter analysis. Theoretically, the population size and maximum iteration positively affect the performance of metaheuristics. Meanwhile, the result in Table III and Table IV indicates that after some level, the increase of maximum iteration or population size remains the same. In the low maximum iteration or population size, the increase of one of these two parameters may improve the result. But, increasing one of these parameters does not produce a better result in the high maximum iteration and population size. In almost all functions, the acceptable solution, whether globally optimal or sub-optimal, has been found in the low maximum iteration and population size.

The third discussion is related to the performance of TOTO in solving the portfolio optimization problem. The result indicates that TOTO performs well in solving this problem by producing the second-best total capital gain. On the other hand, GPA becomes the best one. Fortunately, the performance gap between GPA and TOTO is very narrow. This test also indicates that real-world optimization problems should be used to test all metaheuristics. This result shows that the performance gap in the portfolio optimization problem is narrow compared with the classic 23 functions. It is also indicated that gaining significant improvement in the real-world optimization problem is much more complicated than in the theoretical optimization problem, primarily when the solution is based on an integer number.



The fourth discussion is regarding the limitation of this work and the proposed metaheuristic. This paper has presented that metaheuristics with multiple strategies generally perform better than other metaheuristics with one or two strategies. Meanwhile, TOTO adopts only six strategies (three guided and three random searches). On the other hand, there are a lot of other guided searches, and random searches can be chosen. It means TOTO can be improved by embedding more searches into this metaheuristic. But there are also limitations in embedding new searches. The first limitation is that it can only accommodate some searches into a single metaheuristic. The second limitation is that accommodating various searches into a single metaheuristic is also not wise because some may be less effective than other searches. The question is which searches are better than others. In other words, in which condition is some searches better than others.

The limitation that has been previously discussed makes the development of a new metaheuristic still possible. Although the references used in the guided search converge to the best or randomly selected solution, the selection of these references is still various. For example, many metaheuristics choose the best solution so far like in GPA [21], the best solution in the current iteration like in COA [15], or some best solutions as the leader. In many metaheuristics, for example in RLSBO [27], a randomly selected solution or some randomly selected solutions are uniformly selected among the population. On the hand, like in SO [10], some solutions from some fixed-size groups are randomly chosen to reduce dependency on the best solution that may come with convergence too early.

The second opportunity comes from the motive to minimize the maximum iteration or population size. This work demonstrated that TOTO could perform well in low population sizes and maximum iteration. Theoretically, the performance of any metaheuristic can be improved by increasing the population size or maximum iteration to a very high number. Meanwhile, this work has demonstrated that choosing an appropriate strategy can be a better option than just scaling up the maximum iteration or population size, which is closer to the greedy approach.

The third opportunity comes from the scalability aspect. For example, large-scale problems with very high dimensions can be found easily in many real-world optimization problems, such as optimizing the purchase order of a supermarket with hundreds of stock-keeping units or optimizing the investor's portfolio with hundreds of stocks can be chosen. Large-scale optimization problems consume more computational resources, whether from increasing the maximum iteration or population size so that a sub-optimal solution can be found. It is still challenging to develop a new metaheuristic that does not need excessive computational resources in handling the large-scale problems.

## VI. CONCLUSION

A new swarm-based metaheuristic, namely TOTO, with its superior performance, has been presented in this paper. This paper also has presented a novel mechanism for adopting three guided and three random searches in a competition to find a better solution. The test result indicates that TOTO and precisely the strategy adopted in this metaheuristic perform

better not just in beating some previous metaheuristics but also finding acceptable solutions in the low maximum iteration and low population size circumstances. This work presents that TOTO can find the optimal global solution of eleven functions. Moreover, TOTO is better than GWO, MPA, MLBO, GSO, and GPA in solving 22, 21, 21, 19, and 15 functions, respectively. It means TOTO can tackle problems faced in the three groups of functions. In the test related to a portfolio optimization problem, TOTO is better than GWO, MPA, MLBO, and GSO by producing better total capital gain.

Due to the limitations of this work, this work and especially the proposed metaheuristic can become the baseline for future studies. These future studies can be carried out by improving TOTO or implementing TOTO to solve various real-world optimization problems.

## ACKNOWLEDGMENT

This work was financially supported by Telkom University, Indonesia.

## REFERENCES

- [1] E. Haitam, R. Najat, and J. Abouchabaka, "A vehicle routing problem for the collection of medical samples at home: case study of Morocco", *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 4, pp. 345-351, 2021.
- [2] M. Fekri, M. Heydari, and M. M. Mazdeh, "Two-objective optimization of preventive maintenance orders scheduling as a multi-skilled resource-constrained flow shop problem", *Decision Science Letters*, vol. 12, no. 1, pp. 41-54, 2023.
- [3] T. J. Kumar and M. Thangaraj, "An ordered precedence constrained flow shop scheduling problem with machine specific preventive maintenance", *Journal of Project Management*, vol. 8, no. 1, pp. 45-46, 2023.
- [4] A. Golab, E. S. Gooya, A. A. Falou, and M. Cabon, "A multilayer feed-forward neural network (MLFNN) for the resource-constrained project scheduling problem (RCPSP)", *Decision Science Letters*, vol. 11, no. 4, pp. 407-418, 2022.
- [5] M. Subramanian, M. Narayanan, B. Bhasker, S. Gnanavel, M. H. Rahman, and C. H. P. Reddy, "Hybrid electro search with ant colony optimization algorithm for task scheduling in a sensor cloud environment for agriculture irrigation control system", *Complexity*, vol. 2022, ID. 4525220, pp. 1-15, 2022.
- [6] B. -G. Risi, F. Riganti-Fulginei, and A. Laudani, "Modern techniques for the optimal power flow problem: state of the art", *Energies*, vol. 15, ID. 6387, pp. 1-20, 2022.
- [7] S. R. Singh and R. Chaudhary, "Effect of inflation on EOQ model with multivariate demand and partial backlogging and carbon tax policy", *Journal of Future Sustainability*, vol. 3, no. 1, pp. 35-58, 2023.
- [8] J. Yang, "Big data perspective on financial operations revenue management approach", *Mathematical Problems in Engineering*, vol. 2022, ID. 8901666, pp. 1-9, 2022.
- [9] M. Wang, B. Mao, Y. Yang, R. Shi, and J. Huang, "Determining the level of service scale of public transport system considering the distribution of service quality", *Journal of Advanced Transportation*, vol. 2022, ID. 5120401, pp. 1-14, 2022.
- [10] F. A. Hashim and A. G. Hussein, "Snake optimizer: a novel metaheuristic optimization algorithm", *Knowledge-Based Systems*, vol. 242, ID. 108320, 2022.
- [11] M. Dehghani, S. Hubalovsky, and P. Trojovský, "Tasmanian devil optimization: a new bio-inspired optimization algorithm for solving optimization problem", *IEEE Access*, vol. 10, pp. 19599-19620, 2022.
- [12] M. Dehghani, Z. Montazeri, G. Dhiman, O. P. Malik, R. Morales-Mendez, R. A. Ramirez-Mendoza, A. Dehghani, J. M. Guerrero, and L. Parra-Arroyo, "A spring search algorithm applied to engineering

- optimization problems”, Applied Sciences, vol. 10, ID. 6173, pp. 1-21, 2020.
- [13] S. Arora and S. Singh, “Butterfly optimization algorithm: a novel approach for global optimization”, Soft Computing, vol. 23, no. 3, pp. 715–734, 2019.
- [14] M. S. Braik, “Chameleon swarm algorithm: a bio-inspired optimizer for solving engineering design problems”, Expert Systems with Applications, vol. 174, ID. 114685, pp. 1-25, 2021.
- [15] M. Dehghani, Z. Montazeri, E. Trojovska, and P. Trojovsky, “Coati optimization algorithm: a new bio-inspired metaheuristic algorithm for solving optimization problems”, Knowledge-Based Systems, vol. 259, ID. 110011, pp. 1-43, 2023.
- [16] S. Suyanto, A. A. Ariyanto, and A. F. Ariyanto, “Komodo mlipir algorithm”, Applied Soft Computing, vol. 114, pp. 1–17, 2022.
- [17] M. Dehghani, S. Hubalovsky, and P. Trojovsky, “Northern goshawk optimization: a new swarm-based algorithm for solving optimization problems”, IEEE Access, vol. 9, pp. 162059–162080, 2021.
- [18] S. Z. Koohi, N. A. W. A. Hamid, M. Othman, and G. Ibragimov, “Raccoon optimization algorithm”, IEEE Access, vol. 7, pp. 5383-5399, 2019.
- [19] A. Faramarzi, M. Heidarinejad, S. Mirjalili, and A. H. Gandomi, “Marine predators algorithm: a nature-inspired metaheuristic”, Expert System with Applications, vol. 152, ID. 113377, 2020.
- [20] M. Braik, A. Hammouri, J. Atwan, M. A. Al-Betar, M. A. Awadallah, “White shark optimizer: a novel bio-inspired meta-heuristic algorithm for global optimization problems”, Knowledge-Based Systems, vol. 243, ID. 108457, pp. 1-29, 2022.
- [21] P. D. Kusuma and A. L. Prasasti, “Guided pelican algorithm”, International Journal of Intelligent Engineering and Systems, vol. 15, no. 6, pp. 179-190, 2022.
- [22] S. Kaur, L. K. Awasthi, A. L. Sangal, and G. Dhiman, “Tunicate swarm algorithm: a new bio-inspired based metaheuristic paradigm for global optimization”, Engineering Applications of Artificial Intelligence, vol. 90, ID. 103541, 2020.
- [23] M. I. A. Latiffi, M. R. Yaakub, and I. S. Ahmad, “Flower pollination algorithm for feature selection in tweets sentiment analysis”, International Journal of Advanced Computer Science and Applications, vol. 13, no. 5, pp. 429-436, 2022.
- [24] F. A. Zeidabadi, M. Dehghani, and O. P. Malik, “TIMBO: three influential members based optimizer”, International Journal of Intelligent Engineering and Systems, vol. 14, no. 5, pp. 121-128, 2021.
- [25] F. A. Zeidabadi, S. A. Doumari, M. Dehghani, and O. P. Malik, “MLBO: mixed leader based optimizer for solving optimization problem”, International Journal of Intelligent Engineering and Systems, vol. 14, no. 4, pp. 472–479, 2021.
- [26] M. Dehghani, Z. Montazeri, A. Dehghani, R. A. Ramirez-Mendoza, H. Samet, J. M. Guerrero, and G. Dhiman, “MLO: multi leader optimizer”, International Journal of Intelligent Engineering and Systems, vol. 13, no. 6, pp. 364–373, 2020.
- [27] F. A. Zeidabadi, M. Dehghani, and O. P. Malik, “RSLBO: random selected leader based optimizer”, International Journal of Intelligent Engineering and Systems, vol. 14, no. 5, pp. 529–538, 2021.
- [28] A. Kaveh, S. Talatahari, and N. Khodadadi, “Stochastic paint optimizer: theory and application in civil engineering”, Engineering with Computers, vol. 38, pp. 1921-1952, 2022.
- [29] P. D. Kusuma and D. Adiputra, “Modified social forces algorithm: from pedestrian dynamic to metaheuristic optimization”, International Journal of Intelligent Engineering and Systems, vol. 15, no. 3, pp. 294–303, 2022.
- [30] M. Dehghani, E. Trojovská, and P. Trojovský, “A new human-based metaheuristic algorithm for solving optimization problems on the base of simulation of driving training process”, Scientific Report, vol. 12, No. 1, pp. 1–21, 2022.
- [31] M. Dehghani, S. Hubalovsky, and P. Trojovsky, “A new optimization algorithm based on average and subtraction of the best and worst members of the population for solving various optimization problems”, PeerJ Computer Science, vol. 8, ID: e910, pp. 1-29, 2022.
- [32] M. Noroozi, H. Mohammadi, E. Efatinasab, A. Lashgari, M. Eslami, and B. Khan, “Golden search optimization algorithm”, IEEE Access, vol. 10, pp. 37515–37532, 2022.
- [33] P. D. Kusuma and A. Novianty, “Total interaction algorithm: a metaheuristic in which each agent interacts with all other agents”, International Journal of Intelligent Engineering and Systems, vol. 16, no. 1, pp. 224-234, 2023.
- [34] D. H. Wolpert and W. G. Macready, “No free lunch theorems for optimization”, IEEE Transactions on Evolutionary Computation, vol. 1, no. 1, pp. 67-82, 1997.
- [35] A. Kuznetsov, L. Wieclaw, N. Poluyanenko, L. Hamera, S. Kandy, and Y. Lohachova, “Optimization of simulated annealing algorithm for s-boxes generating”, Sensors, vol. 22, ID. 6073, pp. 1-19, 2022.
- [36] M. A. Noman, M. Alatefi, A. M. Al-Ahmari, and T. Ali, “Tabu search algorithm based on lower bound and exact algorithm solutions for minimizing the makespan in non-identical parallel machines scheduling”, Mathematical Problems in Engineering, vol. 2021, ID. 1856734, pp. 1-9, 2021.
- [37] S. Katoch, S. S. Chauhan, and V. Kumar, “A review on genetic algorithm: past, present, and future”, Multimedia Tools and Applications, vol. 80, pp. 8091-8126, 2021.
- [38] M. Misaghi and M. Yaghoobi, “Improved invasive weed optimization algorithm (IWO) based on chaos theory for optimal design of PID controller”, Journal of Computational Design and Engineering, vol. 6, pp. 284-295, 2019.
- [39] A. G. Gad, “Particle swarm optimization algorithm and its applications: a systematic review”, Archives of Computational Methods in Engineering, vol. 29, pp. 2531-2561, 2022.
- [40] P. Trojovsky and M. Dehghani, “A new optimization algorithm based on mimicking the voting process for leader selection”, PeerJ Computer Science, vol. 8, ID. e976, pp. 1-40, 2022.
- [41] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey wolf optimizer”, Advances in Engineering Software, vol. 69, pp. 46-61, 2014.

# An Empirical Study on the Affecting Factors of Cloud-based ERP System Adoption in Iraqi SMEs

Mohammed G. J<sup>1\*</sup>, MA Burhanuddin<sup>2</sup>, Dawood F.A.A<sup>3</sup>

Alyousif S<sup>4</sup>, Alkhayyat A<sup>5</sup>, Ali M. H<sup>6</sup>, R. Q. Malik<sup>7</sup>, Jaber M. M<sup>8</sup>

Faculty of Information and Communications Technology, Universiti Teknikal Malaysia Melaka, Malaysia<sup>1,2</sup>

Department of Computer Science, University of Baghdad, Baghdad, Iraq<sup>3</sup>

Research Centre, University of Almathreq, Baghdad, Iraq<sup>4</sup>

Department of Electrical and Electronic Engineering, Gulf University, Almasnad, Kingdom of Bahrain<sup>4</sup>

College of Technical Engineering, The Islamic University, Najaf, Iraq<sup>5</sup>

Computer Techniques Engineering Department Imam, Ja'afar Al-sadiq University, Baghdad, Iraq<sup>6</sup>

Medical Instrumentation Techniques Engineering Department, Al-Mustaqbal University College, Babylon, Iraq<sup>7</sup>

Department of Medical Instruments Engineering Techniques, Al-Turath University College, Baghdad, Iraq<sup>8</sup>

Department of Medical Instrumentation Technical Engineer, Al-Farahidi University, Baghdad, Iraq<sup>8</sup>

**Abstract**—This paper aims to investigate the main factors that have an impact on the adoption of cloud-based enterprise resource planning (ERP) among small- and medium-sized enterprises (SMEs) in the Republic of Iraq using TOE, DOI, and HOT-fit as a theoretical framework. Data was collected from 136 decision maker senior executives, and IT managers in SMEs in the Republic of Iraq. A web-based survey questionnaire was used for data collection processes. The research model and the derived hypotheses were tested using SPSS and SmartPLS. The findings indicate several specific factors have a significant effect on the adoption of cloud-based ERP. This conclusion can be utilized in enhancing the strategies for approaching cloud-based ERP by pinpointing the reasons why some SMEs choose to adopt this technology and success during the adoption phase, while others still do not go forward with the adoption. This study provides an overview and empirically shows the main determinants logistical factors that might face SMEs in the Republic of Iraq. The findings also help SMEs consider their information technologies investments when they think to adopt cloud-based ERP.

**Keywords**—SMEs; TOE; DOI and HOT-fit frameworks; Cloud-based ERP; ICT; SmartPLS; SPSS

## I. INTRODUCTION

Business competitiveness can be improved and can provide a significant benefit for businesses when using information and communication technologies (ICTs) [1]. One of the types of businesses are small and medium sized enterprises (SMEs) [2]. SMEs are an important part of every country's economy because they provide most new jobs and drive technical advancement. SMEs account for a more significant proportion of all businesses and GDP than any other sector<sup>1</sup>. SMEs have historically played an essential role in contributing to the economic development of many countries worldwide. SMEs are playing a critical role and considered essential in the world's economies, providing jobs, adding value, and stimulating innovation<sup>1</sup>.

However, SMEs face various logistical challenges and determinants compared to large corporations. They have less

budget allocated and fewer staff [3], making their computing environment less complex than equivalent large enterprises. Moreover, SMEs face requirements similar to those experienced by larger companies. Therefore, SMEs need to improve their service level to achieve the targeted goals of the company. According to [2], SMEs will be expected to modify various logistics to suit the needs of their information technology (IT) departments. Thus, SMEs need to use IT technologies to achieve a higher level of competence and efficiency. The use of enterprise resource planning (ERP) systems offers SMEs numerous tangible benefits. Improving competitive organizational position in the market is considered the most interesting benefit when implementing ERP systems. According to [4], ERP provides a real-time infrastructure for a company's back-end systems, including (purchasing, marketing, sales and inventory, procurement, finance, and human resources) (p.1). However, utilizing standard ERP has several disadvantages for businesses; for instance, on-premise ERP require excessive expenses in terms of initial investments [5]. The author in [6] believed that one of the criteria for SMEs to select ERP systems is to check the affordability. Therefore, there was a need for cloud ERP systems that were lower in cost and less time consuming than traditional ERP systems, and that is where ICT can be an asset.

Due to dramatic business changes, there has been globally widespread use of ICTs among various size enterprises, specifically SMEs. The author in [1] found the use of ICTs by SMEs impacts livelihoods and reduces vulnerabilities. Moreover, applying ICT innovatively in SMEs within most developing countries will increase the opportunity to create jobs. Cloud computing technology is an ICT that helps organizations leverage new IT development at an affordable cost. It has grown in importance as a source of new business service innovations. SMEs will better utilize their limited IT resources by using cloud computing to increase IT system stability and scalability. Therefore, SMEs have been identified as the main beneficiaries of cloud computing [7].

\*Corresponding Author.

Universiti Teknikal Malaysia Melaka (UTeM)

<sup>1</sup><https://www.oecd.org/industry/C-MIN-2017-8-EN.pdf>

In this regard, the field of ERP has seen an evolution through the development of cloud-based ERP systems [2]. Cloud ERP system enables users to transfer and share information in real-time due to its three-layer architecture [4]. Moreover, cloud ERP can be used without IT infrastructure requirements; all updates are reported and processed immediately by the providers [8]. Thereby, cloud-based ERP is a cost-effective alternative to traditional ERPs. As a result, adopting these solutions would be highly beneficial to SMEs.

Cloud-based ERP solutions offer various significant benefits to SMEs due to their cost and time effectiveness. Cloud services enable SMEs to save money by extending the flexibility and agility of internal and external operations while also lowering manufacturing costs. Additionally, the ERP market has expanded with the advent of cloud-based ERP solutions [2]. Cloud ERP systems in past few years are gaining more attention in the market, especially among SMEs, however, it has been recognized that SMEs have low motivation toward the adoption of this technology [9] due to failure of implementation after the adoption of cloud ERP system. The author in [10] stated that SMEs' decision makers face failure after cloud ERP implementation due to lack of knowledge and experience. Therefore, it is critical to identify the factors and recognize the issues leading to implementation failure through the adoption stage. To have a successful adoption, decision makers are required to understand and evaluate the factors affecting the adoption of cloud ERP.

Although several studies have attempted in the past to assist decision-makers in addressing their worries about cloud ERP adoption, it's worth noting that most of these proposed frameworks and models have several limitations. Because present models and frameworks primarily focus on the operational and tactical levels, the frameworks and models themselves are limited (that is, they do not encompass many perspectives). Moreover, there was a lack of research and studies on the factors that affect the use of cloud-based ERP in developing Middle East countries [11]. Additionally, most past studies have been undertaken in developed countries and the North Asia Pacific region [12]. Very few studies have been conducted to study the factors affecting cloud ERP adoption in developing nations, notably among SMEs in the Republic of Iraq. The adoption of cloud ERP systems in the Republic of Iraq is still at an early stage. To the best of the author's knowledge, no previous study has investigated the factors that may influence an SME's decision to adopt a cloud ERP system in the private sector in Iraqi SMEs. Therefore, this study aims to address the abovementioned gap by examining the factors influencing the adoption of cloud ERP in Iraqi SMEs from the perspective of decision-makers. The study addresses the following research questions:

- What are the critical factors leading to the adoption of cloud-based ERP system in Iraqi SMEs?
- How can the relationships between the factors and the adoption of cloud-based ERP can be modelled?

The study integrates three relevant frameworks, which were the Diffusion of innovation (DOI), the technology organization environment (TOE) framework and technology, human, and organization (HOT-fit) framework. Based on

these frameworks and due to the dynamic nature of cloud ERP, a multidisciplinary approach is required to explain the decision-making aspects impacting cloud ERP adoption. This study proposes an integrated model to identify the main factors that will encourage or prevent SME decision-makers in Iraq from moving on to cloud-based ERP systems, using web-based survey system for data gathering. This paper is organised as follows. The literature review of the cloud-based ERP and the problem area following by prior research studies are presented in Section II. In Section III we present our research model and the theoretical frameworks that has been used. The research hypothesis of the research model is presented in Section IV. The Methodology of the study is located in Section V. In Section VI, the findings of our study are presented. Section VII and Section VIII contain the discussion and the contribution. We conclude the paper in Section IX with a conclusion, limitations and future work.

## II. LITERATURE REVIEW

### A. Cloud-based ERP

Cloud computing's popularity has accelerated in recent years [13]. Additionally, the concept of ERP adoption (particularly for on-premise systems) has been extensively investigated. Nevertheless, over the last six years, cloud-based ERP (also known as cloud ERP) has gained popularity [14]. Cloud ERP is defined in this section as software applications that integrate business processes and transaction-oriented data across an organization via a model that enables ubiquitous, convenient, on-demand network access with minimal management effort or service provider response. For decades, ERP systems have been adopted as industry-standard software that satisfies business requirements. With the advent of cloud infrastructure, ERP software vendors are aggressively investing in cloud-based ERP systems, and this deployment model is gaining traction as a practical method of delivering ERP software [15]. Cloud-based ERP systems enable businesses to rent their whole ERP system landscape from various service providers, including software and infrastructure providers.

Future SME growth was significantly correlated with (Internet of Things) implementation. Investments in technology could help SMEs develop more rapidly than previously anticipated. SMEs must use technical innovations and e-commerce products to increase their competitive advantage [2]. With regard to SMEs, similar studies supporting the usage of radio-frequency identification technology to address the idea of technical advancements within SMEs. As a result, technical innovations in relation to the adoption of cloud-ERP systems would be extremely beneficial in accelerating the growth of SMEs.

Cloud-based ERP systems arose in the mid-2000s as an alternative to on-premise ERP systems, particularly for handling upgrades and maintenance operations [16]. Cost pressures on enterprises have led to an increase in the use of cloud computing solutions [17]. The use of cloud services is growing as the cost of these systems falls. Cloud-based ERP system is broadly classified into three categories: SaaS (software-as-a-service), IaaS (Infrastructure-as-a-server) and Hybrid cloud ERP [8].

### B. Prior Research Studies in Cloud ERP Adoption

The use of cloud-based ERP systems by SMEs would contribute to their growth by addressing issues and enhancing their competitiveness [2]. Because cloud ERP would increase productivity and efficiency while also lowering electricity costs, it is a more cost-effective option [10].

Numerous studies have been undertaken to evaluate the benefits and challenges that businesses encounter when implementing cloud ERP in their businesses [9] [18]. Some researchers have concentrated on cloud ERP adoption in various locations such as Europe, the United States, Australia, and Asia [19] [20]. Others have concentrated on the acceptance of cloud ERP in various industries such as banking, healthcare, and manufacturing [11] [21] [22].

The author in [2] investigated the key logistical determinants of cloud ERP adoption in SMES in developing Middle Eastern countries in order to improve cloud ERP approach strategies by examining why some SMEs embrace this technology while others do not. The DOI and TOE theoretical frameworks were used to investigate 14 variables in this study. Although the study contributes to the adoption of cloud ERP, the study model did not include the human context. The author in [23] investigated the difficulties of cloud ERP implementation. They also investigated the primary advantages of issues encountered during cloud-ERP installation. However, their study was not focused on the adoption phase. The author in [24] examined the factors influencing cloud ERP adoption by manufacturing SMEs in Nigeria. The TOE framework and the DOI theory was used to develop a theoretical model and validated the factors using quantitative analysis. Their findings indicated that cost savings, compatibility, and data privacy are the primary and most significant elements influencing the technological background for cloud ERP adoption. However, they ignore decision makers factors. The author in [25] investigated the reasons for SMEs in developing countries' unwillingness to use Cloud ERP. The study has examined several factors that lead to the reluctance of Cloud ERP based on individual, innovation and organizational characteristics from twenty SMEs in Namibia. This study has some limitations regarding the number of data collected. Moreover, the study covered the SMEs during the implementation phase.

The author in [26] conducted a study that incorporated the TOE framework. He tried to ascertain the variations in TOE characteristics between firms that implemented Cloud ERP systems and those that did not. However, the study had no intention of researching any organization's size, and the responders to the results were not decision-makers within the firm. The author in [19] created a theoretical model utilizing the TOE framework and hypothesized the effect of critical factors on the intention of SMEs in Malaysia, to use cloud-based ERP systems. However, their findings have no noticeable effect on cloud-based ERP adoption. Their findings are ambiguous and have limitations when it comes to decision makers' characteristic determinants. The author in [11] combined the TOE and DOI frameworks to construct a model for examining the factors driving cloud ERP system adoption in Saudi Arabian organizations. Their study had limitations that it did not consider critical variables such as security and

privacy, nor did it take into account the characteristics of decision-makers that could influence cloud adoption. The author in [18] explored the challenges to cloud ERP adoption in Pakistani SMEs using a qualitative methodology in association with unstructured interviews. According to the study, ten potential barriers to the adoption were identified the primary issues to determine whether to accept or reject cloud ERP solutions. However, their analysis excluded human features and relied entirely on qualitative data.

Furthermore, while most of previous studies emphasize the technological aspects of cloud ERP, they overlook the human and organizational aspects [12]. Even though some concepts of cloud computing technology are universal, others will differ due to different contexts and country requirements. Cloud ERP adoption also varies by country and industry. Although the researchers identify cloud ERP as a global IT phenomenon, they also emphasize that the factors influencing the adoption of cloud ERP play different roles in different economic environments. Human factors such as the characteristics of decision makers differ from country to country, even if the countries are located in the same region. Despite the importance of cloud-based ERP usage and its role to enhance the company's performance, there was a lack of research and studies on the factors that affect the use of cloud-based ERP in developing Middle East countries [11].

In summary, this study found that the preceding research on cloud-ERP adoption and its factors is twofold. First, the majority of the research has focused on the benefits and limitations of cloud ERP adoption, the development of cloud ERP frameworks, and the variables influencing cloud ERP implementation from a service's perspective in developed countries. However, few studies have been conducted on organizational perspectives, notably those of decision-makers or end-users, which are crucial for business cloud ERP adoption. As a result, numerous research uses the TOE paradigm to investigate the cloud-ERP implementation issue. Secondly, prior studies have emphasized the significance of technological, organizational, and environmental factors in cloud ERP adoption. The adoption of cloud ERP, however, varies depending on the human factors across industrial sectors. For instance, factors that may affect SME adoption decisions made by decision-makers have not been sufficiently examined in conjunction with other aspects in prior studies.

## III. THEORETICAL FRAMEWORK

### A. Research Model

There is a tendency toward merging several frameworks or theories. In order to better explain adoption in the context of cloud-based ERP, the majority of academics combine two or more theories. The complementarity of various existing theories on technology adoption makes this approach applicable. DOI, TOE, and HOT fit theories complement one another to thoroughly understand the major determinants of technologies [22]. The TOE, DOI, and HOT-fit frameworks are the most influential theories in cloud service adoption studies in organizations [22]. The TOE framework is based on the organizational level theory, which provides a multi-perspective framework by incorporating internal and external elements [27]. They also identified three institutional contexts

of the TOE framework, including the technological, organizational, and environmental contexts, which impact the adoption of technological innovations. The author in [28] defined DOI as a process for decreasing uncertainty, and he provided attributes that aid in reducing uncertainties about a technological invention. They are the following: Relative Advantage, Compatibility, Complexity, Trialability, and Observability. The author in [29] presented HOT-fit as an

evaluation paradigm for health information systems. HOT-fit evaluates HIS systems on three dimensions (technology, human, and organization), arguing that user attitude and competence also affect technology adoption.

The present study comprises four dimensions that influence cloud-ERP adoption: Technological context, Organizational contexts, Environmental context, and Human context. Fig. 1 illustrates the research model.

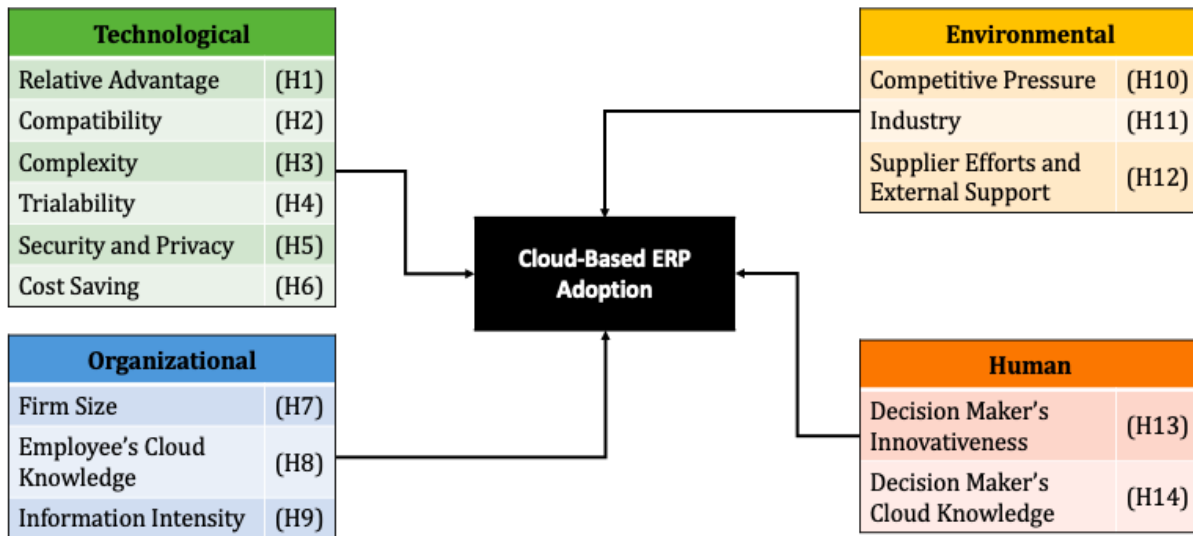


Fig. 1. Research model.

#### IV. RESEARCH HYPOTHESIS

##### A. Technological Context

The technological context in the original TOE framework referred to both internal and external technologies critical to a company [30]. The technological context has been developed by combining the DOI attributes and the technology dimension of TOE framework.

The most influential factors in the innovation adoption have been adopted to create the technological contexts, include: relative advantage, compatibility, complexity, trialability, security and privacy, and cost saving.

The Relative Advantage (RA) factor relates to determining if cloud ERP systems will benefit the indicated SME more than other technologies. This factor is a substantial and favorable influence in the choice to use cloud ERP [13]. According to [31], a company is more likely to adopt cloud ERP in its operations if it recognizes the potential benefits of cloud ERP over other traditional types of ERP. Moreover, RA has been identified as a significant element and is wholly associated with adopting IT innovations [32]. It was discovered that relative advantage has a considerable and beneficial effect on cloud service adoption in Lebanese SMEs [33].

Compatibility (CAT) is defined as the degree to which new technology is compatible with an organization's principles, current demands, and historical traditions. The

author in [9] identified compatibility as a critical aspect in adopting innovation in cloud ERPs. While CAT was identified as a critical factor in the adoption of cloud services by SMEs [34]. However, it was not shown to affect cloud service adoption in Saudi Arabian higher education institutions [35].

Complexity (CEX) is a term that refers to how difficult innovation is to learn, use, and adopt. The more easily a business can integrate with technology, the more rapidly it may be accepted when innovation is primarily intended to add value to the business. Adopting new technology, such as cloud ERP Services, may present obstacles for SMEs, such as the need to modify current processes [36].

Trialability (TRA) has been viewed as a crucial necessary element of adopting new technology [7] [37] [38]. TRA is essential for early adopters than late adopters, as the latter can profit from the previous experience as a benchmark of the innovation's performance.

Security and Privacy (SP) are described as (the extent to which cloud computing is seen as more secure than other computer ideas) [34]. The SP of an organization's data are significant obstacles to shifting data to the cloud. One of the most frequently mentioned concerns concerning cloud-based services is the cloud's security and privacy. The term (cloud security and privacy) relates to various issues, including the cloud's confidentiality and integrity [39].

The Cost Saving (CT) of cloud ERP is defined as (the extent to which decision-makers view the entire cost of employing cloud ERP to be less than the cost of on-premise ERP). Cloud services support innovation by lowering IT investment costs and total IT capital expenditure [33]. The cost was shown to be a significant factor in determining whether to utilize cloud services by Australian SME's [38]. These considerations lead to the following hypotheses:

- H1: Relative advantage (RA) will positively influence the decision to adopt Cloud ERP within SMEs.
- H2: Compatibility (CAT) will positively influence the decision to adopt Cloud ERP within SMEs.
- H3: Complexity (CEX) will negatively influence the decision to adopt Cloud ERP within SMEs.
- H4: Trialability (TRA) will positively influence the decision to adopt Cloud ERP within SMEs.
- H5: Security (SP) and Privacy concerns will positively influence the decision to adopt Cloud ERP within SMEs.
- H6: Cost Saving (CT) will positively influence the decision to adopt Cloud ERP within SMEs.

#### B. Organizational Context

The organizational context relates to the organization's properties and the resources available to foster innovation. In the context of cloud ERP, successful cloud ERP adoption requires the organization to endure both technical and organizational changes. The organizational context in this study includes firm size, employee's cloud knowledge, and information intensity.

The Firm Size (FS) has a significant impact on how quickly a technology innovation is being adopted. Numerous studies have demonstrated that an organization's size can positively influence its adoption of cloud services [7] [37]. The author in [37] demonstrated that the effect of certain variables varied according to organization size.

Numerous studies have discovered that employee's cloud knowledge (ECK) play a significant role in technology adoption decisions [7] [33]. Employees' understanding of similar prior experiences can be viewed as showing the similarity between current practice and prior experience. As a result, an organization's employees' cloud knowledge and experience could be referred to as its technological capability. To use the ERP system through the cloud, staff must acquire both cloud-related expertise and prior ERP system experience.

Information Intensity (INI) is defined as the degree to which a business's product or service contains information. The term (information intensity) relates to the amount of information available within a business service. Businesses in various areas require varying levels of information; for example, financial brokers require access to the most up-to-date information. According to [40], INI positively affects the adoption of cloud services because the cloud can provide up-to-date and appropriate information. According to [41], information intensity increases innovation adoption. Based on

the discussion above, the following hypotheses are proposed for this study:

- H7: Firm size (FS) will positively influence the decision to adopt Cloud ERP within SMEs.
- H8: Employee's knowledge/Experience (ECK) will positively influence the decision to adopt Cloud ERP within SMEs.
- H9: Information Intensity (INI) will positively influence the decision to adopt Cloud ERP within SMEs.

#### C. Environmental Context

The environmental context refers to the location of the firm as indicated by the industry type. The researcher found several factors, including competitive pressure, industry type, and supplier effort and external support that might influence an organization's decision to adopt cloud ERP.

Competitive Pressure (CP) refers to the degree to which an enterprise is under pressure from other external partners, such as competitors within the market [2]. Numerous studies have demonstrated the effect of competitive pressure on technology adoption, such as cloud ERP, in SMEs [26] [42], while other empirical studies have revealed no relationship between competitive pressure and technology adoption [7] [11].

The organization's industry sector (IN) has a significant impact on whether or not it decides to implement new technology [7]. Since firms in different sectors have varying business needs, those that depend on technology are much more willing to adopt a technology invention such as cloud ERP. It has been claimed that the industry in which a business operates affects the information processing requirements of the business, which in turn affects the firm's decision to adopt new technology.

Supplier effort and External support (ES) is described as the amount of assistance in integrating an information system [7]. In the context of cloud ERP, external support refers to assistance provided by cloud server providers, government policies, and legal safeguards to assist SME decision-makers in adopting cloud ERP. Because cloud ERP is seen as new in developing countries, there is a lack of information and existing work on cloud ERP engaging SMEs [11]. Thus, the following hypotheses are proposed:

- H10: Competitive pressure (CP) will positively influence the decision to adopt Cloud ERP within SMEs.
- H11: Industry sector (IN) which SME operate will positively influence the decision to adopt Cloud ERP within SMEs.
- H12: Supplier efforts and external computing support (ES) will positively influence the decision to adopt Cloud ERP within SMEs.

#### D. Human Context

Before adopting any IT project, the human factors should be considered since it affects the successful acceptance of innovative technology [29]. According to [22], the human

context was selected as the most crucial context for adopting cloud services. The following human factors are:

Decision Maker's Innovativeness (DMI) is defined as the degree to which decision-makers are willing to experiment with innovative and perhaps risky solutions [43]. The innovativeness of decision-makers is measured by their willingness and ability to experiment with new technologies that have not yet been thoroughly explored in their industry [44]. Previous studies have found the innovativeness of decision-makers to be a positive factor for cloud services adoption [7] [22] [44]. Studies highlighted that if the individuals are critical decision-makers in the company, personal innovativeness can help identify who will adopt an innovation first [7].

Possessing sufficient knowledge about new technology is the first stage in adopting. According to [45], the CEO's knowledge of information systems (DMC) has a beneficial effect on IS/IT adoption. The technological knowledge and skills of the decision-maker will influence the decision to adopt an IT innovation. SME will have more trust if the decision-maker has proper knowledge and skills on cloud ERP. As a result, the following hypothesis is generated:

- H13: Decision Makers' innovativeness (DMI) will positively influence the decision to adopt Cloud ERP within SMEs.
- H14: Decision Makers' cloud knowledge (DMC) will positively influence the decision to adopt Cloud ERP within SMEs.

## V. METHODOLOGY

### A. Scope, Sampling and Measurement Instrument

The purpose of this study is to investigate the primary determinants and variables impacting a SME's decision makers to adopt cloud ERPs in Republic of Iraq. This study's target respondents were composed of the decision-makers (Owners, CIOs, CEOs, IT executives, and IT managers) belonging to SMEs. The participants who are either (Owners, CIOs, CEOs) or heading departments (IT executives and IT managers), directly involved and responsible for technology acquisition and policy decisions for their SMEs. Managers are presumed to be the decision-makers or contributors to decision-making for adopting new technologies in their enterprises. Although the researchers were able to gain a list of private enterprises from the Ministry of planning website in the Republic of Iraq, several issues were encountered when validating the frame, such as the list not being up to date, the list not specifying the location of the SMEs, a larger number of these enterprises being already closed, inactive numbers, duplication of firms or duplicated phone numbers, and these enterprises not specifying if the enterprises were small or medium-sized. After several iterations and cleaning of this list, 300 unique and validated phone numbers of enterprises were identified, which were considered the core of the sample frame. Each SME in the sample met the following criteria: must be in Baghdad governorate; and must not have more than 30 employees. SMEs in Baghdad government were chosen because they are in a better position to understand current IT

operations, increase in growth rates and future trends for their enterprises [11]. The questionnaire was prepared in English and subsequently translated into Arabic, as the study was conducted in Iraq, where Arabic is the official language. A web-based survey was utilized due to various benefits, including quicker distribution, a more professionalism, affordability, and an improvement in response rate.

Pilot study was necessary to revise the wording, structure, content, arrangement, layout, simplicity, and clarity of the survey instrument [46]. 50 samples from the total sample frame were conducted for the purpose of pilot study. The feedback from the pilot study was necessary to enhance the questions and test respondents' comprehension and clarity before administering the actual survey.

Since the data was analyzed using SEM, this research has determined the sample size according to [47] rule, which indicate the sample size in SEM must equal the larger 10 times the largest number of formative indicators used to measure a single construct. The number of leading indicators utilized for measuring a single construct is 12 for RA; therefore, the sample size is 120 ( $10 \times 12$ ).

Since the sample size of 120 is required for this study, 180 questionnaires were given to respondents to eliminate a questionnaire shortage caused by various factors (missing, ignoring, and incomplete). One hundred fifty-five (155) decision-makers responded to our email and completed the survey, representing an 86.1 percent response rate. Thus, it was acceptable for a web survey that was sent via e-mail/internet [48].

Eleven (11) of 155 respondents were rejected because significant areas of the questionnaires were incomplete. Moreover, eight (8) responses were eliminated since they were ineligible to participate in the survey due to not meeting the criteria. 136 completed surveys were left for the study.

An email was sent to every decision-maker contained a link to the online survey, which was conducted using Google Forms, as well as a cover letter describing the study's aim. The survey contained a definition and description of various terms used in the questionnaire to enhance the validity of the response. A first and second reminder was given to all decision-makers interested in participating in this research study, both by email and separated by a few weeks to ensure the completion of the survey. In total, 75 items for 14 constructs in the survey came from prior studies and literature which was modified by the researcher to fit the context of cloud-based ERP adoption. The survey is divided into two main parts. Part one seeks the demographic information from the respondents; this part also relates demographic information to the respondent's organization as shown in Table I. This part of the questionnaire was related to fundamental questions about the firm, such as whether the SME has adopted on-premise ERP or cloud ERP. Also, questions regarding businesses size, industry, and market scope were included. Part two was about the core constructs developed in the proposed model, which contained the questions used for studying the four contexts, namely technological, organizational, environmental, and human, assessed by five Likert-type items.



TABLE I. DEMOGRAPHIC INFORMATION

Variable	Frequency	Percentage	Variable	Frequency	Percentage
<b>Position</b>			<b>Cloud ERP awareness</b>		
Owner	44	32.4	High awareness	77	56.6
CEO	12	8.8	Low awareness	59	43.4
IT Manager	25	18.4	<b>Firm Size</b>		
IT Executive	35	25.7	1 - 9 Employees	47	34.6
Other	20	14.7	10 - 30 Employees	89	65.4
<b>Gender</b>			<b>Industry Type</b>		
Male	117	86	Technology	31	22.8
Female	19	14	Telecommunications	14	10.3
<b>Age</b>			Energy	9	6.6
20 - 25	18	13.2	Construction	13	9.6
26 - 30	72	52.9	Oil, Gas and Petrochemicals	9	6.6
31 - 35	22	16.2	Manufacturing	12	8.8
36 - 40	14	10.3	Pharmaceutical	6	4.4
41 and above	10	7.4	Education	11	8.1
<b>Education Level</b>			Financial services	7	5.1
Diploma	18	13.2	Private healthcare and services	5	3.7
Bachelor	79	58.1	Entertainment/media	4	2.9
Masters	28	20.6	Transport, Distribution and Logistics	3	2.2
Doctoral	11	8.1	Consumer Packaged Goods	7	5.1
<b>Work Experience</b>			Business trading	5	3.7
0-5	13	9.6	<b>ERP System Type</b>		
6-10	64	47.1	On-premise ERP	44	32.4
11-15	24	17.6	Cloud ERP	32	23.5
16-20	26	19.1	None	60	44.1
+21	9	6.6			

**B. Data Analysis**

The partial least squares (PLS) path modelling approach was used for this work because it is more appropriate for predictive and exploratory research [48]. PLS-SEM is chosen as the most acceptable approach because this study was designed to predict the influencing factors among the specified variables and to analyze their impacts on adoption within the proposed model.

To analyse the data, SmartPLS software was used, which is one of the most well-known software applications for PLS-SEM [49]. Statistical Package for the Social Science (SPSS) software was used to analyze demographic and institutional data, such as industry and market size. In addition, the instrument's reliability was also tested using Cronbach's alpha technique.

**VI. FINDINGS**

**A. Reliability and Validity**

The reliability and validity were measured first. The reliability of data was evaluated using composite reliability and the alpha coefficient. The composite reliability (CR) values ranged from 0.860 to 0.957 and Cronbach's Alpha values ranged from 0.757 to 0.951. Table II shows that all constructs are regarded to have adequate internal consistency reliability were above the acceptable 0.7 threshold.

Validity was measured by considering convergent and discriminant validity. The author in [47] defined convergent validity as the degree of correlation between two measures of the same construct. This study assessed the convergent validity using the AVE [50]. Convergent validity is permitted when AVE is at least 0.5. The study's measuring approach has good convergent validity since all AVE values meet the minimum requirement of 0.50 as shown in Table II [50].

TABLE II. CONSTRUCT CONSISTENCY AND RELIABILITY

Constructs	AVE	CR	CA
AD	0.875	0.934	0.859
RA	0.649	0.957	0.951
CEX	0.747	0.898	0.832
CAT	0.653	0.917	0.888
TRA	0.776	0.932	0.904
SP	0.686	0.945	0.933
CT	0.666	0.952	0.944
IN	0.741	0.895	0.824
ES	0.755	0.939	0.919
CP	0.793	0.920	0.871
ECK	0.831	0.951	0.933
INI	0.720	0.885	0.810
FS	0.786	0.917	0.866
DMI	0.673	0.860	0.757
DMC	0.875	0.934	0.859

The author in [51] defines discriminant validity as the degree of difference between two notions. The discriminant validity of constructs is assessed by the [52] criterion, HTMT and cross-loadings of construct items. The discriminant validity of a construct is supported when the square root of its AVE is greater than its correlation with other constructs [52]. The discriminant validity of a measurement model is determined by the item loadings and the square root of the construct's AVE. The square roots of AVE, shown in Table III, were larger than their corresponding correlation, showing that our data had good discriminant validity.

The HTMT is a new technique for determining discriminative validity in contrast-based SEM. HTMT is the average of all correlations of indicators across combinations that measure different structures relative to each other. The author in [53] suggested that HTMT values below 0.85 indicated distinct combinations. Table IV shows all HTMT values below 0.85. As a result, discriminatory structures were provided.

TABLE III. FORNELL-LARCKER CRITERION

	AD	RA	CEX	CAT	TRA	SP	CT	IN	ES	CP	ECK	INI	FS	DMI	DMC
AD	<b>0.935</b>														
RA	0.799	<b>0.806</b>													
CEX	0.56	0.686	<b>0.864</b>												
CAT	0.599	0.753	0.776	<b>0.808</b>											
TRA	0.515	0.643	0.459	0.622	<b>0.881</b>										
SP	0.691	0.724	0.611	0.732	0.691	<b>0.828</b>									
CT	0.786	0.692	0.687	0.751	0.62	0.775	<b>0.816</b>								
IN	0.461	0.651	0.515	0.597	0.561	0.726	0.626	<b>0.861</b>							
ES	0.62	0.772	0.518	0.625	0.732	0.767	0.735	0.817	<b>0.869</b>						
CP	0.509	0.676	0.478	0.56	0.757	0.731	0.652	0.716	0.796	<b>0.891</b>					
ECK	0.522	0.705	0.615	0.719	0.671	0.732	0.68	0.682	0.701	0.759	<b>0.912</b>				
INI	0.405	0.551	0.481	0.621	0.421	0.557	0.536	0.448	0.401	0.392	0.591	<b>0.849</b>			
FS	0.464	0.515	0.513	0.627	0.396	0.442	0.538	0.445	0.405	0.429	0.42	0.398	<b>0.887</b>		
DMI	0.531	0.606	0.5	0.605	0.687	0.674	0.579	0.603	0.723	0.66	0.63	0.406	0.448	<b>0.82</b>	
DMC	0.796	0.732	0.694	0.745	0.614	0.804	0.596	0.604	0.73	0.626	0.672	0.542	0.537	0.568	<b>0.829</b>

TABLE IV. HTMT

	AD	RA	CEX	CAT	TRA	SP	CT	IN	ES	CP	ECK	INI	FS	DMI	DMC
AD															
RA	0.791														
CEX	0.646	0.757													
CAT	0.679	0.712	0.745												
TRA	0.563	0.676	0.521	0.682											
SP	0.76	0.628	0.681	0.503	0.732										
CT	0.594	0.642	0.762	0.627	0.652	0.788									
IN	0.532	0.734	0.607	0.691	0.637	0.639	0.708								
ES	0.69	0.631	0.588	0.69	0.783	0.729	0.797	0.731							
CP	0.572	0.797	0.554	0.634	0.523	0.797	0.724	0.646	0.779						
ECK	0.566	0.746	0.697	0.785	0.709	0.776	0.725	0.767	0.756	0.749					
INI	0.463	0.611	0.568	0.737	0.485	0.628	0.65	0.542	0.459	0.45	0.672				
FS	0.52	0.563	0.598	0.706	0.425	0.484	0.592	0.519	0.449	0.482	0.454	0.466			
DMI	0.649	0.705	0.621	0.725	0.546	0.706	0.676	0.766	0.665	0.518	0.739	0.505	0.529		
DMC	0.773	0.639	0.777	0.428	0.649	0.657	0.655	0.683	0.793	0.697	0.719	0.611	0.596	0.664	

### B. Hypothesis Testing

Bootstrapping is a statistical approach used in SmartPLS to evaluate the correlations (paths) between dependent and independent variables. The structural model was constructed to identify the path relationship among the variables. The study tests the relationship between endogenous and exogenous variables via the path coefficient ( $\beta$ ) and t-statistics.

The study found that RA ( $\beta=0.100$ ,  $p=0.005$ ,  $t=2.853$ ), CAT ( $\beta=0.106$ ,  $p=0.036$ ,  $t=2.100$ ), TRA ( $\beta=0.138$ ,  $p=0.005$ ,  $t=2.835$ ), SP ( $\beta=0.166$ ,  $p=0.004$ ,  $t=2.893$ ), CT ( $\beta=0.137$ ,  $p=0.001$ ,  $t=3.236$ ), FS ( $\beta=0.152$ ,  $p=0.037$ ,  $t=2.561$ ), ECK ( $\beta=0.105$ ,  $p=0.034$ ,  $t=2.125$ ), IN ( $\beta=0.185$ ,  $p=0.016$ ,  $t=2.411$ ), ES ( $\beta=0.116$ ,  $p=0.025$ ,  $t=2.255$ ), DMI ( $\beta=0.113$ ,  $p=0.032$ ,  $t=2.147$ ), and DMC ( $\beta=0.125$ ,  $p<0.001$ ,  $t=3.611$ ) have a significant impact on the Adoption Decision (AD) of cloud ERP. However, CEX ( $\beta= -0.003$ ,  $p=0.949$ ,  $t=0.064$ ), INI ( $\beta= -0.016$ ,  $p=0.823$ ,  $t=0.224$ ), and CP ( $\beta= -0.034$ ,  $p=0.432$ ,  $t=0.787$ ) have no significant impact on the Adoption Decision (AD) of cloud ERP. Table V indicated the result of the hypothesis.

The T-statistic was used to determine the contribution of independent constructs to the predictor of the dependent construct, AD. The items' t-values ranged from 0.224 to 3.611, reaching the level of significance. All structural model relationships were statistically significant when the p-value  $<0.05$  and the t-value  $>1.96$ . Except for the paths between (CEX, INI, and CP) and AD, where the CEX ( $t=0.064$ ,  $p=0.949$ ), INI ( $t=0.224$ ,  $p=0.823$ ), and CP ( $t=0.787$ ,  $p=0.432$ ), all of which did not meet the recommended value; thus, these paths this path were not significant.

## VII. DISCUSSION

The purpose of the first study question was to identify the factors that influence an SME's decision to adopt a cloud-based ERP system in Iraqi SMEs. This research question was answered based on the literature review related to the adoption of cloud-based ERP. A comprehensive literature review on cloud ERP system, SMEs, trends, strategies, and frameworks for cloud-based ERP system in different sectors was conducted. Moreover, the cloud-based ERP system was critically reviewed the existing and the use of cloud-based ERP system in the SME sector. Literature deeply emphasizes the importance of technological, organizational, environmental, and human characteristics while conducting technology adoption studies. To identify and confirm the critical factors affecting the adoption of the cloud ERP system, data was collected and examined quantitatively. Based on the analysis, eleven factors out of fourteen were found to influence the adoption in SMEs. The findings in each context are discussed as follows:

### A. Technological Context

Regarding RA, CAT, TRA, SP and CT the findings indicate they are significantly influence cloud-ERP adoption in SMEs. The findings are consistent with some previous studies [11] [24] [34] [38] [54]. Conversely, this finding was inconsistent with that obtained by [2]. These results may strongly contribute to the adoption of cloud ERP in SMEs

through utilizing these factors of influence. In regard of CEX, the findings indicate it was not significant to the adoption of cloud-ERP. The result of CEX was consistent with the result from [2].

### B. Organisational Context

In regard to FS and ECK, the results were consistent with several previous findings from cloud ERP adoption studies which report the significant positive influence the adoption [11] [24] [54]. Contrary to expectations, this study did not find a significant relationship between INI and cloud-ERP adoption. The findings of INI were inconsistent with [40], which indicates that cloud services can make the best use of information intensity and thus influence positively.

### C. Environmental Context

Unexpectedly, CP did not significantly affect cloud-ERP adoption. This result was consistent with previous studies [2] [22]. In contrast, [24] and [11] found that CP has a significant impact in association with cloud-ERP adoption.

IN and ES were statistically significant for cloud-ERP adoption in SMEs in Iraq. In terms of IN, the finding of this study is consistent with the finding of some previous studies [7] [54]. However, it is inconsistent with those of [38], which indicates that cloud services can be understood in the context of an overall business's strategy based on agility and responsiveness. In terms of ES, the result is consistent with those of [26] [38] [42], which they found that the factor was positively associated with adoption in SME. In addition, [26] implied that the availability of vendor (external) support could be a vital factor that encourages SMEs' adoption of the cloud ERP system.

### D. Human Context

Interestingly, two predictors in the form of DMI and DMC were found as statistically significant influences in ERP adoption decisions. The findings of DMI and DMC are consistent and similar to the findings of some previous studies [21] [22] [24], which found that the factor was positively associated with adoption in SME.

TABLE V. RESULTS OF HYPOTHESES AND RELATIVE PATHS

Path	$\beta$	t-values	p-values
RA→AD	0.100	2.853	0.005
CAT→AD	0.106	2.100	0.036
CEX→AD	-0.003	0.064	0.949
TRA→AD	0.138	2.835	0.005
SP→AD	0.166	2.893	0.004
CT→AD	0.137	3.236	0.001
FS→AD	0.152	2.561	0.037
ECK→AD	0.105	2.125	0.034
INI→AD	-0.016	0.224	0.823
CP→AD	-0.034	0.787	0.432
IN→AD	0.185	2.411	0.016
ES→AD	0.116	2.255	0.025
DMI→AD	0.113	2.147	0.032
DMC→AD	0.125	3.611	0.000

Following the second study question and to meet the study's objective, the research proposed an integrated model to identify the main factors that could encourage or prevent SME decision-makers from moving on to cloud-based ERP systems. The initial model integrates the critical factors from the literature review as the main factors that may impact the SME's intention to adopt cloud-based ERP. This research integrates three theoretical frameworks (DOI theory, TOE framework, and HOF-fit framework) to develop the conceptual research model. To refine and confirm the factors in the cloud-based ERP adoption model, the study utilized quantitative methods. A questionnaire survey was used to test the proposed model and confirm the identified factors for the cloud-based ERP adoption. Eleven of the fourteen hypotheses were eventually supported. Both study questions were answered, and study objectives were achieved.

## VIII. CONTRIBUTION

### A. Theoretical Contribution

Based on a review of the literature and to the best of the researchers' knowledge, this study is the first exploratory study that combines DOI, TOE and HOF-fit in identifying the determinants that will affect the adoption of cloud ERP in a developing Middle Eastern nation, particularly, the Republic of Iraq. It aimed to determine the significant link between technological, organizational, environmental, and human context; and the adoption decision of cloud-based ERP in Iraqi SMEs. The effect of several factors, such as relative advantage, compatibility, complexity, trialability, security and privacy, cost saving, firm size, employee's cloud knowledge, information intensity, supplier efforts and external computing support, competitive pressure, industry, decision maker's innovativeness, and decision maker's cloud knowledge towards cloud ERP adoption is an important topic that should be considered further in the future. The results also provide further support for the utility of the TOE, DOI and HOF-fit in technology adoption such as cloud ERP.

### B. Practical Contribution

For SMEs' decision-maker and cloud service providers, the outcomes of this study can be effectively applied. Cloud service providers can use the research model to improve their knowledge about why certain SMEs decide to adopt new technologies while others in the SME sector do not, since cloud service providers need to be aware of typical issues experienced in SMEs when deciding on a new adoption of cloud services. This will be a guide to assess the cloud-based ERP system, providing the SMEs' decision-makers the opportunity to trial this solution before the actual implementation and thus improve the awareness of cloud services. Moreover, SMEs (decision-makers) will evaluate the cloud services (cloud-based ERP systems) before their actual use, allowing them to check the level of compatibility and complexity with their existing system. This study exposes that the decision-makers of some enterprises in the private sector still lack enough knowledge to make an informed choice. Therefore, the decision to adopt this technology depends not only on the decision-maker's lack of innovation, and the adoption requires both innovativeness and knowledge to adopt technological innovations. As a result, the outcomes of this

study will give practical instructions for the effective adoption of cloud ERP in Iraq, as well as help other emerging economies in similar situations prepare for and implement cloud ERP services.

### C. Academic Contribution

Compared with other models of the adoption decision study, the proposed model is intended to be more comprehensive where it falls in the adoption decision phase of the cloud ERP life cycle by assessing the decision to adopt the system by identifying the factors that influence the decision to adopt cloud ERP. This study will open opportunities for more research and enhance the constructs further to clarify the adoption of cloud-based ERP systems in SMEs. The results confirmed that the proposed model fitted the situation well with the data. Therefore, the developed model is seen as valid and can contribute significantly to explaining the adoption of the cloud ERP system. Additionally, the study contributes to the Iraqi innovations technology and enterprises literature by identifying the significant variables that affect SME adoption of the cloud-based ERP systems. A long-ignored decision-makers characteristic is introduced that integrates variables pertinent to a human context.

## IX. CONCLUSION

Enterprises facing business issues have demonstrated a significant desire for cloud ERP. Yet, research on the variables impacting the adoption of cloud ERP, particularly in the growing Middle Eastern nations, is scarce. This report sheds light on the variables that SMEs' decision makers must consider before adopting cloud ERP, with a focus on the challenges faced by decision makers of Iraqi SMEs from different business scopes. The literature review served as the foundation for this study's integrated conceptual model, which was then developed and put to the test, utilizing quantitative research. Therefore, it is anticipated that this research will advance theory, methodology, and practice. The researchers studied the body of literature on the adoption of cloud ERP, the TOE framework, the theory of innovation characteristics (DOI) and HOF-fit framework. The study also provided an in-depth analysis of the present relevant theories and indicated the initial (technological, organizational, environmental, and human) influences on the adoption of cloud ERP. The TOE framework, DOI theory and HOF-fit framework were specifically cited as the best suitable theories for creating a conceptual model for SMEs adopting cloud ERP. Eleven factors (RA, CAT, TRA, CT, SP, FS, ECK, IN, ES, DMI, AND DMC) were found to be significant determinants factors of cloud-ERP adoption. Three factors (CEX, INI, and CP) were insignificant. The findings of this study will enhance the adoption of cloud ERP in the Republic of Iraq and other developing middle Eastern countries. The cloud vendors and SMEs' decision maker will benefit from the findings in overcoming obstacles and effectively adopt cloud ERP.

This study endeavoured to investigate the impact of the most notable factors in the adoption of cloud ERP rather than including a complete list of all possible influencing factors. Future studies could examine at more variables and see how they affect the adoption of cloud ERP. Second, this study was conducted in the Republic of Iraq, the findings of which

cannot represent the overall situation of a developing country. Moreover, data were obtained in a specific governorate, SMEs in Baghdad, limiting the generalizability of the findings. Further research should be done to collect data from different developing countries and additional research could expand the scope of the study by examining another governorate within the Republic of Iraq. Third, the decision makers answered questions based on their observations, work experiences, and ICT understanding. Thus, the data collected may not be sufficiently objective. However, due to the nature of this study (exploratory research), the quality of the collected data is acceptable. Finally, the sample selection was quite limited and must be expanded in future work.

#### ACKNOWLEDGMENT

The authors would like to thank BIOCORE Research Group, Center of Advanced Computing Technology (C-ACT), Fakulti Teknologi Maklumat dan Komunikasi (FTMK) and Centre for Research and Innovation Management (CRIM), Universiti Teknikal Malaysia Melaka (UTeM) for providing the facilities and support for this research.

#### REFERENCES

- [1] S. Chege and D. Wang, "Information technology innovation and its impact on job creation by SMEs in developing countries: an analysis of the literature review," *Technology Analysis and Strategic Management*, vol. 32, no. 3, pp. 256-271, 2020.
- [2] M. AL-Shboul, "Towards better understanding of determinants logistical factors in SMEs for cloud ERP adoption in developing economies," *Business Process Management Journal*, vol. 25, no. 5, pp. 887-907, 2019.
- [3] U. Usman, M. Ahmad, N. Zakaria and A. Alkurdi, "A review of key factors of cloud enterprise resource planning (ERP) adoption by SMEs," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 16, pp. 3884-3901, 2017.
- [4] K. Surendro, "Academic Cloud ERP Quality Assessment Model," *International Journal of Electrical and Computer Engineering*, vol. 6, no. 3, pp. 1038-1047, 2016.
- [5] R. Pareek, "Analytical study of cloud ERP and ERP," *International Journal of Engineering and Computer Science*, vol. 3, no. 10, pp. 8710-8714, 2014.
- [6] S. Venkatraman and K. Fahd, "Challenges and success factors of ERP systems in Australian SMEs," *Systems*, vol. 4, no. 2, p. 20, 2016.
- [7] Y. Alshamaila, S. Papagiannidis and F. Li, "Cloud computing adoption by SMEs in the north east of England: A multi-perspective framework," *Journal of Enterprise Information Management*, vol. 26, no. 3, pp. 250-275, 2013.
- [8] M. Abd Elmonem, E. Nasr and M. Geith, "Benefits and challenges of cloud ERP systems – A systematic literature review," *Future Computing and Informatics Journal*, vol. 1, no. 1-2, pp. 1-9, 2016.
- [9] K. Salum and M. Abd Rozan, "Exploring the challenge impacted SMEs to adopt cloud ERP," *Indian Journal of Science and Technology*, vol. 9, no. 45, pp. 1-8, 2016.
- [10] N. Alsharari, M. Al-Shboul and S. Alteneiji, "Implementation of cloud ERP in the SME: evidence from UAE," *Journal of Small Business and Enterprise Development*, 2020.
- [11] A. AlBar and M. Hoque, "Factors affecting cloud ERP adoption in Saudi Arabia: An empirical study," *Information Development*, vol. 35, no. 1, pp. 150-164, 2019.
- [12] S. Yasiukovich and M. Haddara, "Tracing the clouds. A research taxonomy of cloud-ERP in SMEs," *Scandinavian Journal of Information Systems*, vol. 32, no. 2, pp. 237-285, 2020.
- [13] P. Hsu, S. Ray and Y. Li-Hsieh, "Examining cloud computing adoption intention, pricing mechanism, and deployment model," *International Journal of Information Management*, vol. 34, no. 4, pp. 474-488, 2014.
- [14] S. Salim, "Moving from Evaluation to Trial: The Case of Cloud ERP Adoption in SMEs," University of Technology Sydney, Sydney, 2015.
- [15] T. Sindane, "Developing a Framework to Understand the Challenges and Benefits of Cloud-Based ERP Systems in the South African Mining Industry," University of Pretoria, Pretoria, 2017.
- [16] E. Bjelland and M. Haddara, "Evolution of ERP Systems in the Cloud: A Study on System Updates," *Systems*, vol. 6, no. 2, p. 22, 2018.
- [17] D. Atukwase, "An Analysis of Use of Cloud Enterprise Resource Planning Systems in South Africa," Nelson Mandela Metropolitan University, 2015.
- [18] M. Awan, N. Ullah, S. Ali, I. Abbasi, M. Hassan, H. Khattak and J. Huang, "An Empirical Investigation of the Challenges of Cloud-Based ERP Adoption in Pakistani SMEs," *Scientific Programming*, 2021.
- [19] L. Qian, A. Baharudin and A. Kanaan-Jebna, "Factors affecting the adoption of enterprise resource planning (ERP) on cloud among small and medium enterprises (SMES) in Penang, Malaysia," *Journal of Theoretical and Applied Information Technology*, vol. 88, no. 3, pp. 398-409, 2016.
- [20] J. Rodrigues, P. Ruivo, B. Johansson and T. Oliveira, "Factors for adopting ERP as SaaS amongst SMEs: The customers vs. Vendor point of view," *Information Resources Management Journal (IRMJ)*, vol. 29, no. 4, pp. 1-16, 2016.
- [21] F. Alharbi, A. Atkins and C. Stanier, "Understanding the determinants of Cloud Computing adoption in Saudi healthcare organisations," *Complex & Intelligent Systems*, vol. 2, no. 3, pp. 155-171, 2016.
- [22] S. Almubarak, "Factors Influencing the Adoption of Cloud Computing by Saudi University Hospitals," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 1, pp. 41-48, 2017.
- [23] B. Signe, B. Dace and S. Edgars, "Cloud based cross-system integration for small and medium sized enterprises," *Procedia Computer Science*, vol. 104, pp. 127-132, 2017.
- [24] U. Usman, M. Ahmad and N. Zakaria, "The determinants of adoption of cloud-based ERP of Nigerian's SMEs manufacturing sector using TOE framework and DOI theory," *International Journal of Enterprise Information Systems (IJEIS)*, vol. 15, no. 3, pp. 27-43, 2019.
- [25] V. Hasheela, K. Smolander and T. Mufeti, "An investigation of factors leading to the reluctance of SaaS ERP adoption in Namibian SMEs," *The African Journal of Information Systems*, vol. 8, no. 4, p. 1, 2016.
- [26] J. Kinuthia, "Technological, organizational, and environmental factors affecting the adoption of cloud enterprise resource planning (ERP) systems," Eastern Michigan University, 2014.
- [27] L. Tornatzky, M. Fleischer and A. Chakrabarti, *Processes of technological innovation*, Mass: Lexington books, 1990.
- [28] E. Rogers, *Diffusion of Innovations*, New York: Free Press, 2003.
- [29] M. Yusof, J. Kuljis, A. Papazafeiropoulou and L. Stergioulas, "An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit)," *International Journal of Medical Informatics*, vol. 77, no. 6, pp. 386-398, 2008.
- [30] F. Cruz-Jesus, A. Pinheiro and T. Oliveira, "Understanding CRM adoption stages: empirical analysis building on the TOE framework," *Computers in Industry*, vol. 109, pp. 1-13, 2019.
- [31] A. Razzaq and A. Mohammed, "Cloud ERP in Malaysia: Benefits, challenges, and opportunities," *International Journal*, vol. 9, no. 5, 2020.
- [32] S. S. Alam, Z. Zain, M. Ahmad and M. H. Ali, "Adoption of Cloud Computing by SMEs in Malaysia: Empirical Study," in *Proceedings of the 8th International Conference of The Asian Academy of Applied Business*, Sabah, Malaysia, December 2017.
- [33] M. Skafi, M. Yunis and A. Zekri, "Factors influencing SMEs' adoption of cloud computing services in Lebanon: An empirical analysis using TOE and contextual theory," *IEEE*, vol. 8, no. 2020, pp. 79169-79181, 2020.
- [34] R. Sandu, E. Gide and S. Karim, "The impact of innovative strategies to influence the adoption of cloud based service success in Indian small and medium enterprises (SMEs)," *International Journal of Arts & Sciences*, vol. 10, no. 2, pp. 403-413, 2017.

- [35] A. Tashkandi and I. Al-Jabri, "Cloud computing adoption by higher education institutions in Saudi Arabia: an exploratory study," *Cluster Computing*, vol. 18, no. 4, pp. 1527-1537, 2015.
- [36] D. Barakah, A. Alrobia and S. Alwakeel, "Strategic plan and development projects for modern health clinical information systems at King Saud Medical City," *International Journal of Information and Electronics Engineering*, vol. 4, no. 4, p. 317, 2014.
- [37] N. Alkhater, R. Walters and G. Wills, "An empirical study of factors influencing cloud adoption among private sector organisations," *Telematics and Informatics*, vol. 35, no. 1, pp. 38-54, 2018.
- [38] S. Alismaili, M. Li, J. Shen, P. Huang, Q. He and W. Zhan, "Organisational-level assessment of cloud computing adoption: Evidence from the Australian SMEs," *Journal of Global Information Management (JGIM)*, vol. 28, no. 2, pp. 73-89, 2020.
- [39] Y. Abdulsalam and M. Hedabou, "Security and Privacy in Cloud Computing: Technical Review," *Future Internet*, vol. 14, no. 1, p. 11, 2021.
- [40] O. Ali, J. Soar, J. Yong and H. McClymont, "Exploratory Study to Investigate the Factors Influencing the Adoption of Cloud Computing in Australian Regional Municipal Governments," *Journal of Art Media and Technology*, vol. 1, no. 1, pp. 1-13, 2017.
- [41] M. Jaganathan, R. Mahmood, S. Ahmad and I. Ahmad, "Effect of environmental context on ICT adoption among rural-based small and medium enterprises in Malaysia," *Advances in Environmental Biology*, vol. 8, no. 9, pp. 563-569, 2014.
- [42] F. Safari, N. Safari, A. Hasanzadeh and A. Ghatari, "Factors affecting the adoption of cloud computing in small and medium enterprises," *International Journal of Business Information Systems*, vol. 20, no. 1, pp. 116-137, 2015.
- [43] R. El-Haddadeh, "Digital innovation dynamics influence on organisational adoption: the case of cloud computing services," *Information Systems Frontiers*, vol. 22, no. 4, pp. 985-999, 2020.
- [44] S. Tehrani and F. Shirazi, "Factors influencing the adoption of cloud computing by small and medium size enterprises (SMEs)," in *In Human Interface and the Management of Information. Information and Knowledge in Applications and Services*, Cham, Springer, June 2014, pp. 631-642.
- [45] A. Khayer, M. Talukder, Y. Bao and M. Hossain, "Cloud computing adoption and its impact on SMEs' performance for cloud supported operations: A dual-stage analytical approach," *Technology in Society*, vol. 60, 2020.
- [46] M. Saunders, P. Lewis and A. Thornhill, *Research Methods for Business Students*, Harlow: Pearson Education, 2019.
- [47] J. F. Hair, G. Hult, C. Ringle and M. Sarstedt, *A primer on partial least squares structural equation modeling (PLS-SEM)*, Thousand Oaks: : Sage Publications, 2016.
- [48] K. Kyri, S. Stephenson and J. Langley, "Assessment of nonresponse bias in an internet survey of alcohol use," *Alcoholism: Clinical and Experimental Research*, vol. 28, pp. 630-634, 2004.
- [49] J. F. Hair, M. Sarstedt, C. Ringle and S. Gudergan, *Advanced issues in partial least squares structural equation modeling*, Thousand Oaks: Sage Publications, 2017.
- [50] M. Janadari, S. Sri Ramalu, C. Wei and O. Abdullah, "Evaluation of measurement and structural model of the reflective model constructs in PLS-SEM," in *Proceedings of the 6th International Symposium-2016 South Eastern University of Sri Lanka (SEUSL)*, Oluvil, Sri Lanka, December 2016.
- [51] N. Urbach and F. Ahlemann, "Structural equation modeling in information systems research using partial least squares," *Journal of Information technology theory and application*, vol. 11, no. 2, pp. 5-40, 2010.
- [52] C. Fornell and D. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," *Journal of Marketing Research*, vol. 18, no. 1, pp. 39-50, 1981.
- [53] J. F. Hair, W. Black, B. Babin and R. Anderson, *Multivariate Data Analysis*, London: Cengage Learning EMEA, 2019.
- [54] I. Senarathna, C. Wilkin, M. Warren, W. Yeoh and S. Salzman, "Factors that influence adoption of cloud computing: An empirical study of Australian SMEs," *Australasian Journal of Information Systems*, vol. 22, 2018.

# A Learning-based Correlated Graph Model for Spinal Cord Injury Prediction from Magnetic Resonance Spinal Images

P.R.S.S.V Raju<sup>1</sup>, Dr. V. Asanambigai<sup>2</sup>, Dr. Suresh Babu Mudunuri<sup>3</sup>  
Research Scholar: CSE, Annamalai University, Annamalainagar, India<sup>1</sup>  
Assistant Professor: CSE, Annamalai University, Annamalainagar, India<sup>2</sup>  
Associate Professor: IT, SRKREC, Bhimavaram, India<sup>3</sup>

**Abstract**—In epidemiological research on spine surgery, machine learning represents a promising new area. It is made up of several algorithms that work together to identify patterns in the data. Machine learning provides many benefits over traditional regression techniques, including a lower necessity for a priori predictor information and a higher capacity for managing huge datasets. Recent research has made significant progress toward using machine learning more effectively in spinal cord injury (SCI). Machine learning algorithms are employed to analyze non-traumatic and traumatic spinal cord injuries. Non-traumatic spinal cord injuries often reflect degenerative spine conditions that cause spinal cord compression, such as degenerative cervical myelopathy. This article proposes a novel correlated graph model (CGM) that adopts correlated learning to predict various outcomes published in traumatic and non-traumatic SCI. In the studies mentioned, machine learning is used for several purposes, including imaging analysis and epidemiological data set prediction. We discuss how these clinical predictive models are based on machine learning compared to traditional statistical prediction models. Finally, we outline the actions that must be taken in the future for machine learning to be a more prevalent statistical analysis method in SCI.

**Keywords**—Spinal cord injury; regression; machine learning; graph model

## I. INTRODUCTION

Movement and sensory impulses from the spinal cord, peripheral nerve system, and brain (SC) are key conducts. The nervous system is made of SC and the brain. It has a tubular structure and grey and white matter, including spinal tracks (the bodies of neurons) [1]. SCI is due to the damage in the spinal tracks while carrying information, and damage in the motor and nervous systems results in [2]. Patients may experience paralysis or have their organs cease working properly due to an SCI. We can evaluate SCI patients more precisely because of the motor and sensory ratings provided by the International Standards for Neurological Classification of Spinal Cord Injury (ISNCSCI). American Spinal Injury Association (ASIA) created these scores, which have since been modified multiple times [3]. They are crucial for determining a patient's SCI sufferer's prognosis in a therapeutic rehabilitation program since they are connected with functional status [4]. For a reliable diagnosis of SCIs, clinical evaluation based on ISNCSCI scores has limitations.

It relies on the patient's input, which is subjective and ambiguous when there is concurrent damage to other organs [5].

For the diagnosis of SCI, conventional MRI is frequently employed. A medical imaging technique called MRI creates detailed macroscopic images of organs and tissues. It uses a striking image (black & white) to discriminate between hard and soft tissues [6]. Modern technology called Diffusion Tensor Imaging (DTI) employs echo-planar MRI data. Using the tissue's architecture and structure, it monitors the movement via the SC and brain tissue, water molecules [7]. DTI is used to study pathological conditions and disorders such as multiple sclerosis, hypertensive encephalopathy, and brain tumours. It gives quantitative data on the size and placement of a three-dimensional (3D) space containing each tissue. Diffusion anisotropy is the word used for this. Numerous floating diffusion ellipsoids make up the diffusion tensor [8]. Each diffusion ellipsoid's orientation is specified by a group of vectors that indicate orientation, often called eigenvectors. A distinct outcome matching an eigenvalue is produced when an eigenvector's length or direction is altered. DTI allows for the expression of diffusion anisotropy as fractional anisotropy (FA) [9]. FA, which has a scale from 0 to 1, is frequently used to determine the degree of fibre integrity since it is sensitive to the number of directionally directed fibres per voxel. Water diffusion anisotropy is measured by the FA value, with a higher degree indicated by a higher FA value [10]. In this research, we provide a brand-new Machine learning-based SC analysis technique. One of the main professionals in diagnosing SCI is [11]. Currently, choices are made using human specialists' analysis of FA values and DTI pictures. If we can provide them with more factual data, they will be able to diagnose more precisely. Classification systems are used in machine learning to make predictions or diagnose problems [12] – [13]. However, classification jobs call for training data. We create a training dataset for our method utilizing four FA values from patient and healthy control slice images. The base dataset is then expanded to 15 features present in a dataset with more dimensions, and the intended dataset is abstracted to increase classification precision [14]. Prediction accuracy for the generated dataset is higher than 90%. Our two contributions are using a classification technique to predict SCIs and creating a training set of images produced by the DTI. The data connected to a specific person

is more than 200 MB [15]. There is a huge challenging factor to employ in computer-aided diagnosis; raw data is large. The major research challenge is the complexity in accurate prediction of spinal cord injury using least sample dataset. Some existing machine learning approaches fail to give better accuracy due to lesser number of samples. However, this can be resolved using the advanced learning approaches. This motivates to adopt a novel learning approach for predicting accurate spinal cord injury and to enhance the accuracy with available samples. In our plan, we take the raw DTI data and extract meaningful numerical information that we subsequently use for diagnosis. Any field in which DTI is used for diagnosis can use our method. This work intends to validate the efficiency of the anticipated CGM and explores the prediction ability with the construction of SCI-based functional connectivity. Here, the behavioural relationship between the injury regions is analyzed from the available online dataset. The proposed CGM constructs the connectivity pattern among the injured region to predict the differences from other regions. The experimental outcomes demonstrate that the anticipated CGM-based prediction model outperforms the overall approach significantly. The model is more reliable and stronger in its prediction nature.

The work is organized as follows: Section II offers a comprehensive analysis of prevailing approaches; Section III gives a detailed analysis of the proposed graph-based model and correlation analysis. In Section IV, the numerical analysis of the anticipated model is provided, and the results are discussed. The summary is provided in Section V.

## II. RELATED WORKS

Machine learning is a broad field that primarily applies computing models to many real-world situations. The primary objective of machine learning relies on the development of algorithms using information from a database. Recognition, diagnostics, planning, robot control, and prediction are tasks that can be accomplished when it is employed. Additionally, it can use machine learning to analyze neuroimaging data and predict tissue toxicity [16]. Both tasks use pattern recognition which requires the identification of numerous important variables. Nowadays, enormous data is employed by the researchers that must be managed, analyzed and used. Large amounts of data may conceal significant linkages and correlations that are uncovered via machine learning. It enhances the effectiveness of systems and machine design. Medical image analysis, lesion segmentation, and computer-aided diagnosis have turned as key application areas for machine learning.

In biological studies, classification is a major task where machine learning is crucial to the classification process. With well-known dataset, unknown sample data can be predicted using machine learning. It can distinguish two or more disparate items, combine related objects, or divide different objects. During the classification process, objects are categorized based on their unique characteristics and each item is given a class name to indicate the specific category to which it belongs ("patient," "normal") [17]. Predictions are made using training and testing of unlabelled data. Test dataset contains unidentified sample that are required to

establish the class label and the performance of the training dataset is evaluated. Two popular categorization methods are k-NN and SVM). K-NN is based on instances of feature-space classifiers that select the most nearby data points for classification choices.

The numbers of characteristics are redundant and irrelevant while classification accuracy is preserved by feature selection. Statistical ML is widely utilized before classification and creates a powerful and stable predictor. In addition to noisy data, feature selection manages exceedingly big datasets. Feature selection makes classification faster and efficient by reducing the dimensions of the data. Feature selection methods like Clearness-Based Feature Selection (CBFS), Features selection based on a distance discriminant (FSDD), R-value-based Feature Selection (RFS), ReliefF, and CBFS are some examples of feature selection. R-value [18], is a statistic for measuring the region of overlap between classes in a feature. Identifying traits that promote effective class separability across classes and maximizing the proximity of samples within the same classes form the basis of the FSDD method. ReliefF is one of the most effective feature selection methods. The concept is to estimate feature weights iteratively based on how well they can distinguish between nearby examples. Based on "CScore" metrics, CBFS is an excellent feature selection technique. Many samples were located in the right class region was determined by the score presents an alternative feature selection approach based on the Lasso. This method establishes a scoring system to determine the "quality" of each distinct feature. Several samples are created using training data and then high-relevance feature orderings are chosen for each sample. Finally, highly relevant properties are integrated. This study use selection of features to assess each feature's discriminative power and identify the most distinct feature subset [19] – [20].

Sagittal and axial panels have undergone T1- and T2-weighted imaging to assess SCI separately. Clinical evaluation is performed to gauge neurological damage and its seriousness was measured using MRI technology. The signal change level and clinical outcomes were linked [21]. We automatically applied classification technique to distinguish patient image slices. The system quickly and accurately produces outcomes and integrates algorithms easily. A key indicator of a prediction quality is classification accuracy. Many academics have tried to increase classification accuracy through algorithm or dataset improvements. Obtaining FA values from DTI is to help people find the impacted area. Human experts heavily rely on the personal knowledge they have gained from earlier assessments of T1- or T2-weighted pictures, even though FA value validates SCI. An automated system can identify SCI to diagnose the condition of the affected area and offer pertinent data would be beneficial [22].

Since there is currently no cure for SCI, individuals with motor-related injuries have little chance of sustaining voluntary movement recovery over the long term (more than a year after the injury) [23]. There is growing evidence that neuro-modulation may be viable for chronic and persistent SCI based on recent reports of effective partial functional recovery. [24]. Even with these positive case studies and series, there are still a lot of problems to be fixed before a



conclusive clinical trial. The ability to customize each patient's spatial and temporal neuro-modulation has increased with the technological advancements in implantable neuro-modulation platforms. However, large, high dimensional flexibility necessitates effective algorithmic optimization customized to each patient's unique pathologies and underlying physiologic system.

Based on simulation conditions, research using animal models has shown huge responses in voluntary movement. Further evidence of this variance was found in human models, necessitating the spatiotemporal adjustment of eSCS-customized patient-specific characteristics [25]. It is crucial to find the optimal parameters because the heterogeneity of SCIs may cause the observed variation in response to the stimulation parameters. A reliable system for choosing the best settings must be created for electrical stimulation therapy because there are trillions of potential configurations for 16-lead paddle. Some studies are currently available on eSCS optimization techniques for causing volitional movement after cSCI. The majority of research uses animal models, and various teams make use of various optimization strategies. To identify the spinal circuits and fibres drawn by eSCS, the author developed computer model that merged 3D finite element approach with rat spinal cord model. The ideal parameters for standing and walking rats are predicted using different electrode configurations. Specific muscle responses to be evaluated by electromyography (EMG) in spinally transected rats were chosen by bipolar stimuli based on Bayesian optimization [26]. Another animal experiment focused on enhancing stepping by adjusting stimulus intensity, the time between pulses and strength. Some staff members determined the ideal frequency and intensity subjectively. Then, kinematic data, EMG, and various stimulation pulse intervals were used to re-evaluate these parameters. To find the best parameter combinations, one study used 3D kinematic data recorders and quantitative gait characteristics [27] – [28].

Even less research has been done on enhancing stimulation in human model systems than in animal models. Various optimization goals and techniques have been tested in various investigations. A map of each participant's motor neuron activation was produced to identify where the spinal cord was engaged during particular muscle movements. Using computerized model, the best electrode combinations were identified through simulations [29]. EMG mapping data has also chosen [30]. The limitations of current optimization approaches make it evident that all-encompassing strategy for choosing the best model for SCI prediction. The major research gap is a lack of proper methodology for feature selection and classification even with small datasets. This may leads to poor prediction outcomes. Therefore, this research concentrates on modelling an efficient approach for prediction.

### III. METHODOLOGY

This section gives a detailed explanation of the proposed model for soft tissue prediction using learning concepts. We initially provide a brief overview of a few terminologies, and there are some definitions of graphs and graph signals. Using SVR, it is then determined how the FC patterns relate to the

appropriate behavioural measure; refer to the framework in Fig. 1. Next, we build FC patterns using the spinal image and our correlated graph model (CGM). Finally, we use simulated data to validate the proposed CGM.

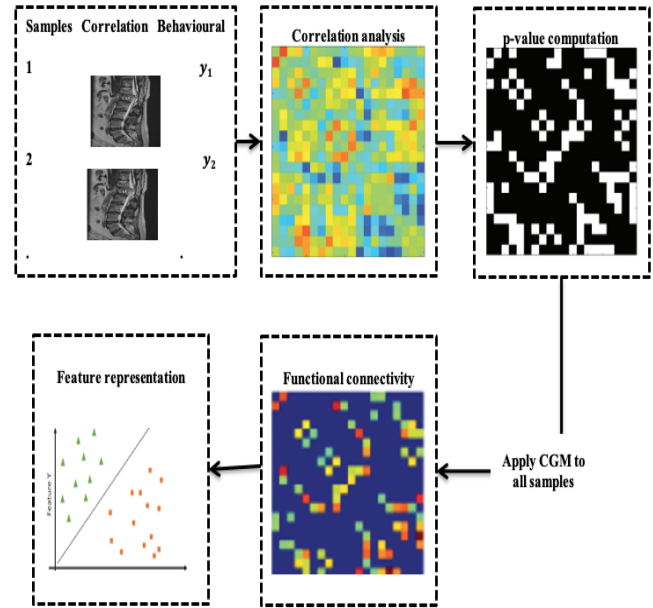


Fig. 1. Framework on a predictive model.

#### A. Graph model

Consider the FCN to be an undirected, linked, weighted graph. Let's build a graph with  $G = \{V, W\}$  where  $V = \{v_1, v_2, v_N\}$  is a collection of  $N$  nodes (ROIs) and  $W = [W_{ij}]_{i,j=1}^N \in R^{N \times N}$  is a weighted adjacency matrix that is symmetric (and frequently sparse), with  $W_{ij} \geq 0$  representing the degree of similarity between the nodes. The degree matrix generated by the diagonal matrix with its diagonal  $D \in R^{N \times N}$ ,  $L \in R^{N \times N}$  member  $D_{ii} = \sum_{j=1}^N W_{ij}$  is referred to as the "Laplacian matrix."

$$L = D - W \quad (1)$$

As a result, we use either the Laplacian matrix,  $L$  or the weighted adjacency matrix of graph  $W$ , which may uniquely describe the underlying graph, to define the  $FC$  pattern quantitatively. With  $f = [f_1, f_2, f_3, \dots, f_N]^T \in R^N$ ,  $f$  be a signal on the graph  $G$  that links value-based feature selection and features selected using a distance discriminant, overall variability about the Laplacian matrix  $L$  as depicted in Eq. (2):

$$f^T L f = \frac{1}{2} \sum_{i \neq j} W_{ij} (f_i - f_j)^2 \quad (2)$$

The smoothness of the graph-wide variation of a graph signal measures the size of the change. Given that nodes with high edge weights are densely coupled, it makes sense that when  $W_{ij}$  is sizable, the gap between  $f_i$  and  $f_j$  will also be narrow. As a result, many machine learning methods like graph regularization and transductive learning have successfully used this graph smoothness notation. According to the perspective of graph signal processing, by utilizing the

graph frequencies provided by the eigenvalues to define different degrees of graph signal smoothness, the eigenvectors of  $L$  offer Fourier transform for graph signals. The  $FC$  pattern in this work was different from the  $FC$  pattern predicted by our CGM due to the magnitude of squared spectral coherences among the time courses of paired ROIs and the examination of brain signals related to diverse graph frequencies.

### B. Correlation Analysis

Consider that  $M$  training subjects are present. In the image courses, let  $X^{(i)} = [x_1^{(i)}, x_2^{(i)}, \dots, x_N^{(i)}]^T \in R^{N \times K}$ . Assume there are  $M$  training participants with  $i^{th}$  subject  $1 \leq i \leq M$  where  $x_j^{(i)} \in R^k$  specifies the ROI in the  $i^{th}$  subject and  $N$  represents the number of ROI. Let  $y_i, 1 \leq i, = M$  specify the subjects' measure. Consider  $X^{(i)}$  is normalized to pose zero mean and unit normalization. The Pearson's correlation matrices  $R^{(i)} \in R^{N \times N}, 1 \leq i \leq M$  for all individuals,  $R_{ijk}^{(i)}$  specifies correlation coefficient among  $j^{th}$  and  $k^{th}$  ROI in the  $i^{th}$  individual, i.e.

$$R_{jk}^{(i)} = corr(x_j^{(i)}, x_k^{(i)}) \quad (3)$$

The behavioural measure is then coupled with Pearson correlation matrices; correlation matrices  $Q \in R^{N \times N}$  and  $P \in R^{N \times N}$  correspond to each between-ROI correlation coefficient.

$$Q_{jk} = corr(r_{jk}, y) \quad (4)$$

Where  $h$  refers correlation coefficient across all people,  $r_{jk} = [R_{jk}^{(1)}, R_{jk}^{(2)}, \dots, R_{jk}^{(M)}]^T \in R^M$  is the  $(j, k)^{th}$  among ROI correlation coefficient and  $y = [y_1, y_2, \dots, y_M]^T \in R^M$  is the behavioural measure. Then, we choose the significant correlation coefficients in  $Q$  from the  $p$ -value matrix  $P$  that are associated with the behavioural measure using a certain threshold. Thus, the matrix  $C \in R^{N \times N}$  is expressed as:

$$C_{jk} = \begin{cases} 1 & \text{if } P_{jk} < \tau \\ 0 & \text{else} \end{cases} \quad (5)$$

With the label information of the subjects, we can drastically reduce the amount of redundant or irrelevant  $FC$  characteristics by employing matrix  $C$  as a guide for learning the  $FC$  pattern. To estimate the  $FC$  pattern, the CGM technique, in addition to the conventional following, is the graph learning technique:

$$\min_{\tilde{L} \in R^{N \times N}} tr \left( (X^{(i)})^T (\tilde{L}^{(i)} \odot C) X^{(i)} \right) + \beta \left\| \tilde{L}^{(i)} \odot C \right\|_F^2 \quad (6)$$

$$tr(\tilde{L}^{(i)} \odot C) = N \quad (7)$$

$$(\tilde{L}^{(i)} \odot C)_{jk} = \left( (\tilde{L}^{(i)} \odot C)_{kj} \right) \leq 0, j \neq k \quad (8)$$

$$\left( (\tilde{L}^{(i)} \odot C) \right) \cdot I_N = 0_N \quad (9)$$

Here,  $L(i) = \tilde{L}^{(i)} \odot C$  and  $C$  refer graph's Laplacian matrix ( $FC$  pattern), where  $C$  is computed with Eq. (5),  $\beta > 0$  refers to the positive regularization parameter and  $\|\cdot\|$  and  $\odot$  specifies matrix norm and element-wise product respectively. The CGM transforms into the conventional graph

learning methodology when the threshold is 1 or there is no correlation guidance, as in Eq. (6), where the matrix  $C = 1_{N \times N}$ .

It is important to note that by minimizing their fluctuations on the learnt graph, the first component of the goal of Eq. (6) is to match the observed spinal cord injury image with the learned graph. The second term helps to further eliminate duplicate  $FC$  characteristics by regulating the sparsity of  $L^{(i)}$  as denser as is greater and vice versa. Additionally, the second and third conditions are included to guarantee that the first restriction is imposed as normalization and learned  $L^{(i)}$  refers to a legitimate positive semidefinite Laplacian matrix. There is a safeguard against trivial solutions. Cross-validation will be utilized to identify the hyper-parameter  $\beta > 0$ .

To create subject-specific  $FC$  patterns that reflect a common template matrix  $C$ , the CGM was constructed in Eq. (6) for both the link between ROIs and the graph organization between ROIs. Specifically, by obtaining the  $1 \leq (i) \leq M$  learned Laplacian  $L^{(i)}$  matrices, we use graph-weighted matrices  $W(i)$  for  $1 \leq (i) \leq M$  to reflect  $FC$  patterns of  $M$  individuals by applying the convex optimization. The vector length  $N(N-1)/2$  represents the  $FC$  features of the  $i$ -th individual produced by triangle  $W$ 's symmetric section.

From these results, the CGM successfully recovers  $FC$  patterns that have more power to discriminate between people, as they are intrinsically relevant to the targeted behavioural measure suggesting the superior performance of behavioural prediction. Therefore, we do a regression analysis to determine how the generated  $FC$  patterns relate to each other using linear SVR with default settings and the behavioural measure in this work to verify the efficacy of the suggested CGM. Fig. 1 is used to train a prediction model. To further clarify how subjects are split into training and test sets, see 1 for more information. After the test subjects' generated  $FC$  patterns are created, the training set's framework is fed into the trained predictive model to provide projected behavioural measures. Although there are many other regression models, the higher effectiveness of the recommended CGM for the optimal  $FC$  pattern prediction is the focus of this study rather than the best regression model schemes.

Concerning the simulated data ( $K = 100$ ) on  $N = 12$  node random weighted network, we verified the CGM made two processes to create the random graph. First, the graph's structure was created with 0.5 probability connections between each pair of nodes, giving the linked edge between the two nodes a uniformly distributed random weight between 0 and 1. Then, a random multiplication of the  $C$  matrix of the  $N \times N$  zero-one template. With edge weights, the graph-weighted matrix  $W$  was produced by linearly coupling the first four eigenvectors of Laplacian matrix  $L$ . Here,  $K = 100$  graph signals are then produced as linear combinations of the eigenvectors (first four) of the graph Laplacian matrix  $L$ .

$$x = 3 \sum_{i=1}^4 \alpha_i f_i \quad (10)$$

Where  $f_i$  refers to an  $i^{th}$  eigenvector of  $L$  and  $\alpha_i$  refers uniform random variable with  $[0, 1]$  range. We inferred the graph structure from these graph signals regarding the

different values of  $\beta$ , specifically the values of  $\beta \in \{0.5, 1, 5, 10, 50, \text{ and } 80\}$ . We discovered that  $\beta = 10$  produced the highest level of performance as determined by the graph Laplacian matrix and Normalized Mutual Information (NMI) discovered during signal processing and the ground truth. The learned graph Laplacian matrices are shown in Fig. 2 for  $\beta = 1, 10, \text{ and } 80$ , respectively. We used a separate zero-one template matrix  $C_b$ , which had 40 randomly generated  $C$  elements to emphasize the significance of ensuring a CGM-acceptable zero-one template matrix. Comparing the NMI performance of the CGM with  $C$  and  $\hat{C}$ , we varied, as shown in Table I.

TABLE I. TEMPLATE MATRIX

	$\beta = 0.5$	$\beta = 1$	$\beta = 5$	$\beta = 10$	$\beta = 50$	$\beta = 80$
$C$	0.50	0.37	0.61	0.71	0.51	0.48
$\hat{C}$	0.23	0.23	0.23	0.15	0.048	0.03

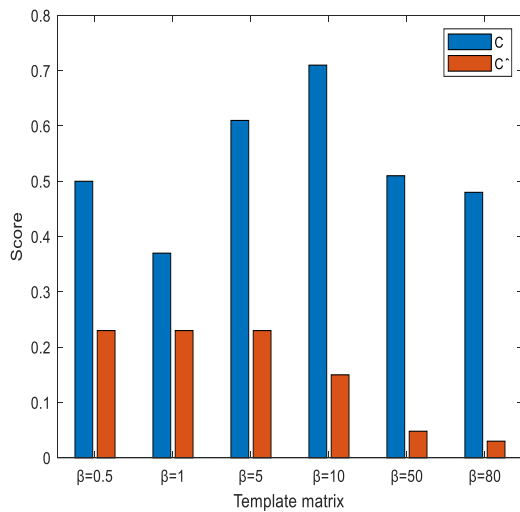


Fig. 2. Matrix evaluation.

#### IV. EXPERIMENTAL SETTINGS

In our trials, we used state SCI data to predict various measures. Using a 5-fold CV, five times using SVR was used to predict individuals' behaviour using FC patterns derived from the CGM (with LIBSVM). The entire group of patients was divided into five roughly equal-sized disjoint subsets by chance; chose one subset at random as the test set, and the training set was used with the other four. 20 rounds of this process were performed to reduce how sample bias affects cross-validation (CV). The prediction performance (average) over all 20 iterations was reported using a 5-fold CV. The hyper-parameter  $\beta$  was found by employing the grid search ranges from  $[0.05, 2.5]$  and the inner CV of the training set. The correlation between the anticipated and observed behavioural variables among people in the test set and Root Mean Square Error (RMSE) were used to evaluate the prediction performance. The template matrix  $C$  was generated at a random significance level. To find the optimal threshold, with values of 0.01, 0.05, 0.1, 0.15, and 0.25, we repeatedly examined five distinct  $p$ -value criteria.

#### A. Dataset

The significance of MRI is to provide superior discriminate soft tissues, along with its capability to acquire heterogeneity and tumour changes. The available online dataset known as Cancer Imaging Archive is used in this research. The dataset comprises CT/PET/MRI scans of 51 patients. Another dataset with 21 patients is also considered. Here, 11 pathologically verified Liposarcomas arise with the soft tissue, and 10 Leiomyosarcomas influence muscle cells. The cohort is composed of 9 females and 12 males with a duration of 31 months. Ground truth discriminating histopathological subtypes is definite. Tumours are localized in the pelvis, biceps, and thigh. Here, three different types of MRI are utilized for training T1-, T2-weighted fat-saturated, and short tau inversion recovery (STIR). With the T1 sequences, the data acquisition is made of the axial plane, while STIR and T2FS are acquired in diverse orientations (coronal, sagittal, and axial). MRI scan with slice thickness is 5.5 mm for T1 and 5 mm for T2-weighted fat saturated. The plane resolution is  $0.63mm^2, 0.74mm^2, \text{ and } 0.86mm^2$  for T2FS, T1, and STIR scans.

#### B. Result Analysis

It's important to remember that in the proposed CGM, the graph model is used to evaluate and identify functional linkages that we need to assess. A significance test is then performed on the degree to which these chosen linkages are connected and the between ROI correlation coefficients with the participant-wide behavioural measure of interest as in Fig. 3. Each behavioural measure shows the mean SCI functional network patterns among patients employing the related techniques separately. The PC-based FC patterns are clear to see based on graph learning and are significantly denser than the FC patterns based on the graph model. The proposed model generates FC patterns that are substantially less dense than those generated by the correlated graph model. The CGM-based FC patterns show greater variability in multiple functional connections, further reinforcing connectivity with high functionality strength and vice versa. It highlights CGM's usefulness and combines pair-wise correlation and graph learning. It assesses the prediction performance of the suggested model for each behavioural indicator in Table II and Table III. The recommended CGM had the best  $cc$  and  $RMSE$ -based prediction performance. By contrasting these outcomes, we may demonstrate the suggested CGM's efficacy and efficiency based on its superior performance. To extract more discriminative FC patterns for building SCI-behavior linkages, it may be advantageous to merge temporal correlation among ROIs with the graph structure across ROIs (GL) of CGM.

We also looked into how the parameters in the suggested CGM affected the accuracy of behavioural prediction. First, the  $p$ -value threshold used in the overall results of the CGM for each behavioural measure is impacted by the method used to construct the template matrix. The  $p$ -value threshold determines how many functional connections in the CGM must be learned. With a lower  $p$ -value threshold, The CGM's graph learning stage will accept fewer connections, indicating a more rigorous connection selection process. Except for

WRAT prediction, the  $p$ -value threshold selection affects how well other predictions. Second, we directly computed the prediction results using a fixed and carefully chosen  $p$ -value cutoff for creating the template matrix and 20 iterations of the 5-fold cross-validation with various hyper-parameter values. It was done to investigate the CGM's sensitivity to the relevant regularization parameter  $\beta$ . Fig. 4 presents the  $cc$  results. We discovered that the results varied depending on the regularization parameter's value and obtained the best performance in each case.

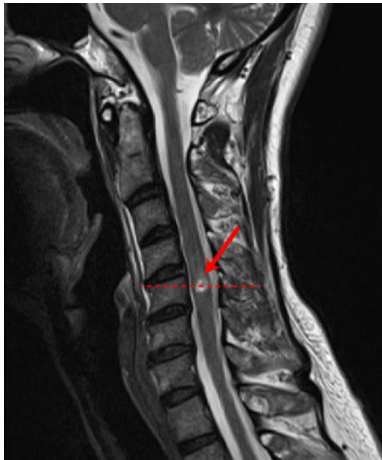


Fig. 3. ROI of SCI.

TABLE II. CC AND RMSE COMPARISON

Methods	Behaviour	$cc(\text{mean} \pm SD)$	RMSE ( $\text{mean} \pm SD$ )
CGM	Injured	<b><math>0.63 \pm 0.013</math></b>	<b><math>3.12 \pm 0.03</math></b>
Type II Fuzzy with CNN+VGG-16		$0.16 \pm 0.03$	$0.16 \pm 0.03$
DRNN		$0.36 \pm 0.02$	$2.11 \pm 0.02$
NB		$0.23 \pm 0.03$	$2.16 \pm 0.03$
k-NN		$0.38 \pm 0.03$	$14.67 \pm 0.03$
PCA+NB		$0.33 \pm 0.03$	$15.11 \pm 0.03$
PCA + k-NN		$0.28 \pm 0.03$	$15.16 \pm 0.03$
PCA + SVM		$0.56 \pm 0.03$	$13.05 \pm 0.03$

TABLE III. ACCURACY DETECTION

Methods	Without Inclusion	With Inclusion	Average
CGM	<b>99%</b>	<b>100%</b>	<b>99.5%</b>
Type II Fuzzy with CNN+VGG-16	96.8%	100%	99%
DRNN	96%	99%	98%
NB	93%	100%	98%
k-NN	93%	100%	98%
PCA+NB	86%	78%	80%
PCA + k-NN	90%	93%	92%
PCA + SVM	86%	100%	96.5%

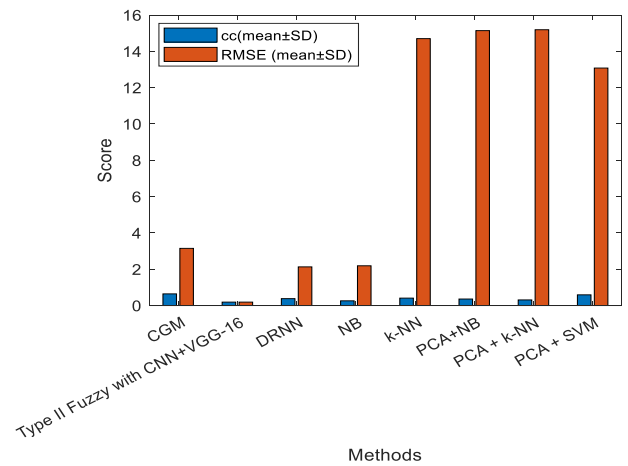


Fig. 4. CC and RMSE comparison.

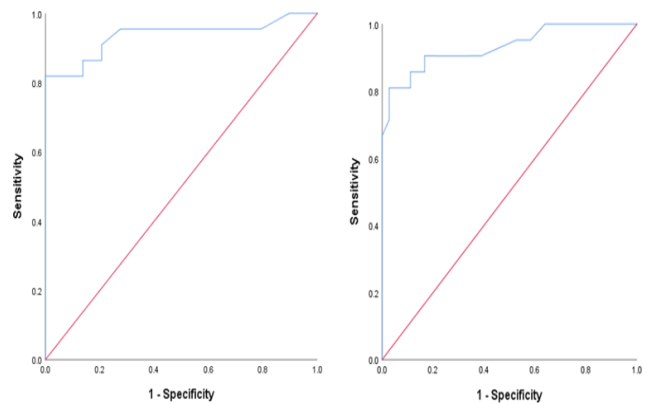


Fig. 5. Validation of non-traumatic SCI.

Finally, we evaluated the most distinctive biologically significant functional relationships that may be linked to these three behavioural characteristics using the FC patterns produced by the proposed CGM. As imaging biomarkers affect a person's variance in the mentioned behavioural assessments, the identified connections may be a supervised learning technique called the SVR assigns distinct attributes distinct weights to most closely mimic the response values in the training set. In this investigation, we emphasized the linkages to which the trained SVR gave more weight. Finally, using the FC patterns produced by the proposed CGM, we independently explored the most discriminative functional links with hypothesized biological values associated with these three behavioural traits. Imaging biomarkers, which influence the individual variation, the relationships found may be employed in the aforementioned behavioural tests. Fig. 5 and Fig. 6 depict the performance comparison of the proposed model.

### C. Discussion and Analysis

In this study, we presented the CGM, which builds by considering the graph structure between ROIs, the relationship between ROIs at different times, more discriminating FC patterns, and constructing brain-behaviour correlations that have been discovered. The CGGL was then used to individually predict three behavioural measures of SCI data from the resting state using the publically available PNC

datasets. In prediction performance, this method outperformed competing FC pattern estimating techniques. The CGM offers a potentially reliable and efficient solution for examining the connections between the brain and behaviour to estimate FC patterns.

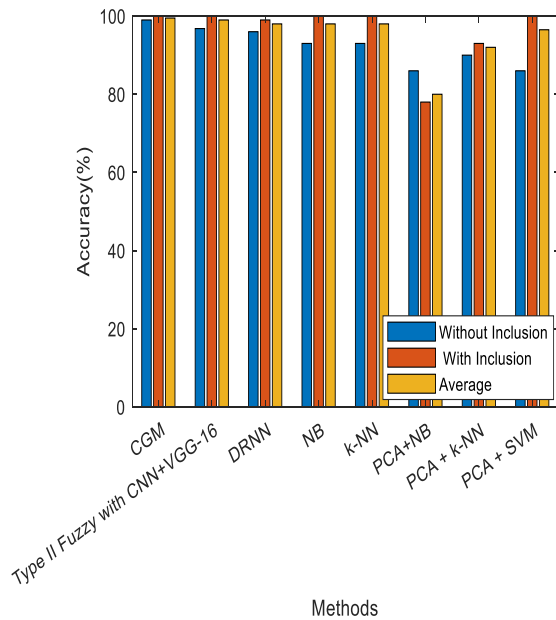


Fig. 6. Accuracy comparison.

The FC pattern generation methods based on deep learning have drawn more attention recently, largely because of the incredibly high prediction performance. Fundamentally, layer-by-layer learning from SCI courses is the basis for deep learning-based FC patterns, which typically have more complicated hidden information. On the other hand, deep learning techniques inevitably entail several variables. Substantial training datasets are frequently required to account for the weights and biases of the various layers and expensive computing power to optimize these parameters. However, with fewer parameters, our suggested technique can produce discriminative FC patterns. In the case of small samples, this reduces the over-fitting issue and increases generalization capacity and effectiveness. Gaining an anatomical understanding of which functional relationships result in individual variation in the relevant behavioural measure was another goal of this investigation.

## V. CONCLUSION

In this study, we introduced the CGM, a novel technique for creating spinal cord injury FC pattern patterns. The recommended CGM combines two widely utilized FC pattern analyses with graph learning and Pearson's correlation. Both the graph structure across ROIs and the relationship between ROI points and time were considered. As a result, the suggested CGM has a lot of promise for improving the generated FC patterns' prediction ability for establishing SC injury correlations and collecting insightful knowledge about the biological processes involved in the behavioural measures of interest. By independently predicting three behavioural variables, we assessed the effectiveness and efficiency of our

suggested CGM using available data from the available sources. The experiment's findings supported the proposed CGM's superiority over other FC pattern estimating techniques which have broad implications in brain network analysis. In future, this study can be further extended with adoption of a novel optimization approach for attaining the global outcomes in terms of accuracy and prediction. Also, with the adoption of pre-trained model, the time complexity can be reduced effectually.

## REFERENCES

- [1] Kepler CK, Vaccaro AR, Koerner JD, Dvorak MF, "Reliability analysis of the AOSpine thoracolumbar spine injury classification system by a worldwide group of naïve spinal surgeons", *Eur Spine J.* 2016;25(4):1082-6.
- [2] Sarrami P, Armstrong E, Naylor JM, Harris IA. Factors predicting outcome in whiplash injury: a systematic meta-review of prognostic factors. *J Orthop Traumatol.* 2017;18:9-16.
- [3] Boban J, Thurnher MM, Van Goethem JW. Spine and spinal cord trauma. In: Barkhof F, et al., editors. *Clinical Neuroradiology.* Cham: Springer; 2018.
- [4] Talbott JF, Whetstone WD, Readdy WJ, "The brain and spinal injury centre score: a novel, simple, and reproducible method for assessing the severity of acute cervical spinal cord injury with axial T2-weighted MRI findings", *J Neurosurg Spine.* 2015;23(4):495-504.
- [5] Rutman AM, Vranic JE, Mossa-Basha M. Imaging and managing blunt cerebrovascular injury. *Radiographics.* 2018;38:542-63.
- [6] Poplawski MM, Alizadeh M, Oleson CV, Fisher J, Marino RJ, Gorniak RJ, Leiby BE, Flanders AE. Application of diffusion tensor imaging in forecasting neurological injury and recovery after human cervical spinal cord injury. *J Neurotrauma.* 2019;36:3051.
- [7] Shanmuganathan K, Zhuo J, Chen HH, Aarabi B, Adams J, Miller C, Menakar J, Gullapalli RP, Mirvis SE. The diffusion tensor imaging parameter obtained during acute blunt cervical spinal cord injury predicts long-term outcomes. *J Neurotrauma.* 2017;34:2964-71.
- [8] Scivoletto, G., Tamburella, F., Laurenza, L., Torre, M., & Molinari, M. (2014). Who is going to walk? A review of the factors influencing walking recovery after spinal cord injury. *Frontiers in human neuroscience*, 8, 141.
- [9] Saadeh, Y. S., Smith, B. W., Joseph, J. R., Jaffer, S. Y., Buckingham, M. J., Oppenlander, M. E., ... & Park, P. (2017). The impact of blood pressure management after spinal cord injury: a systematic review of the literature. *Neurosurgical Focus*, 43(5), E20.
- [10] Sabit, B., Zeiler, F. A., & Berrington, N. (2017). The impact of mean arterial pressure on functional outcome post-acute spinal cord injury: a systematic scoping review of animal models. *Journal of neurotrauma*, 34(18), 2583-2594.
- [11] Yue, J. K., Hemmerle, D. D., Winkler, E. A., Thomas, L. H., Fernandez, X. D., Kyritsis & Talbott, J. F. (2020). Clinical implementation of novel spinal cord perfusion pressure protocol in acute traumatic spinal cord injury at US Level I Trauma Center: TRACK-SCI study. *World Neurosurgery*, 133, e391-e396.
- [12] Squair, J. W., Bélanger, L. M., Tsang, A., Ritchie, L., Mac-Thiong, J. M., Parent, S., & Street, J. (2019). Empirical targets for acute hemodynamic management of individuals with spinal cord injury. *Neurology*, 93(12), e1205-e1211.
- [13] Gaudin, X. P., Wochna, J. C., Wolff, T. W., Pugh, S. M., Pandya, U. B., Spalding, M., & Narayan, K. K. (2019). Incidence of intraoperative hypotension in acute traumatic spinal cord injury and associated factors, *Journal of Neurosurgery: Spine* SPI, 32(1), 127-132.
- [14] Catapano, J. S., Hawryluk, G. W. J., Whetstone, W., Saigal, R., Ferguson, A., Talbott, J., & Manley, G. (2016). Higher mean arterial pressure values correlate with neurologic improvement in patients with initially complete spinal cord injuries. *World neurosurgery*, 96, 72-79.
- [15] Hawryluk, G., Whetstone, W., Saigal, R., Ferguson, A., Talbott, J., Bresnahan, J., & Manley, G. (2015). Mean arterial blood pressure correlates with neurological recovery after human spinal cord injury:

- analysis of high-frequency physiologic data. *Journal of neurotrauma*, 32(24), 1958-1967.
- [16] Altaf, F., Griesdale, D. E., Belanger, L., Ritchie, L., Markez, J., Ailon, T., Boyd, M.C., Paquette, S., Fisher C.G., Street, J. & Dvorak, M. F. (2017). The differential effects of norepinephrine and dopamine on cerebrospinal fluid pressure and spinal cord perfusion pressure after acute human spinal cord injury. *Spinal Cord*, 55(1), 33-38.
- [17] Bao, F. P., Zhang, H. G., & Zhu, S. M. (2017). Anaesthetic considerations for patients with acute cervical spinal cord injury. *Neural regeneration research*, 12(3), 499.
- [18] Krishna, V., Andrews, H., Varma, A., Mintzer, J., Kindy, M. S., & Guest, J. (2014). Spinal cord injury: how can we improve the classification and quantification of its severity and prognosis? *Journal of neurotrauma*, 31(3), 215-227.
- [19] Sherubha, "Graph-Based Event Measurement for Analyzing Distributed Anomalies in Sensor Networks", *Sādhanā*(Springer), 45:212, 2020, <https://doi.org/10.1007/s12046-020-01451-w>.
- [20] Sherubha, "An Efficient Network Threat Detection and Classification Method using ANP-MVPS Algorithm in Wireless Sensor Networks", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, ISSN: 2278-3075, Volume-8 Issue-11, September 2019.
- [21] Sherubha, "An Efficient Intrusion Detection and Authentication Mechanism for Detecting Clone Attack in Wireless Sensor Networks", *Journal of Advanced Research in Dynamical and Control Systems (JARDCS)*, Volume 11, issue 5, Pg No. 55-68, 2019.
- [22] Kurpad, S., Martin, A. R., Tetreault, L. A., Fischer, D. J., Skelly, A. C., Mikulis, D., & Fehlings, "Impact of baseline magnetic resonance imaging on neurologic, functional, and safety outcomes in patients with acute traumatic spinal cord injury", *Global spine journal*, vol. 3 (1), 2017.
- [23] Czyz M, Tykocki T, Szewczyk P, Jarmundowicz W. Application of diffuse tensor imaging in the outcome prognosis after cervical spinal cord injury. *J Spinal study surgery* 2017.
- [24] Martin, A. R., Aleksanderek, I., Cohen-Adad, J, "Translating state-of-the-art spinal cord MRI techniques to clinical use: a systematic review of clinical studies utilizing DTI, MT, MWF, MRS, and fMRI", *NeuroImage: Clinical*, 10, 192-238, 2016.
- [25] Dvorak, M. F., Noonan, V. K., Fallah, N., Fisher, C. G., Finkelstein, J., Kwon, B. K., ... & Townson, A. (2015). The influence of time from injury to surgery on motor recovery and length of hospital stay in acute traumatic spinal cord injury: an observational Canadian cohort study. *Journal of neurotrauma*, 32(9), 645-654.
- [26] Wengel, P. V., De Haan, Y., Feller, R. E., Oner, F. C., & Vandertop, W. P. (2019). Complete Traumatic Spinal Cord Injury: Current Insights Regarding Timing of Surgery and Level of Injury. *Global Spine Journal*.
- [27] Mac-Thiong, J. M., Li, A., Ehrmann Feldman, D., Gagnon, D. H., Thompson, C., & Parent, S. (2016). Do patients with complete spinal cord injury benefit from early surgical decompression? Analysis of neurological improvement in a prospective cohort study. *Journal of neurotrauma*, 33(3), 301-306.
- [28] Wengel, P. V., de Witt Hamer, P. C., Pauptit, J. C., van der Gaag, N. A., Oner, F. C., & Vandertop, W. P. (2019). Early surgical decompression improves neurological outcome after complete traumatic cervical spinal cord injury: a meta-analysis. *Journal of neurotrauma*, 36(6), 835-844.
- [29] Hidalgo, J. L. T., Ruiz-Picazo, D., Martin-Benlloch, A., TorresLozano, P., & Portero-Martinez, E. (2018). The impact of the urgent intervention on the neurologic recovery in patients with thoracolumbar fractures. *Journal of Spine Surgery*, 4(2), 388.
- [30] Li, J., Wang, H., Yang, Q., Lv, D., Zhang, W & Ma, K. (2014). Posterior short segment pedicle screw fixation and TLIF to treat unstable thoracolumbar/lumbar fracture. *BMC musculoskeletal disorders*, 15(1), 40.

# The Practices of Online Assessment in a Digital Device in the Context of University Training: The Case of Hassan II University

## How to Evaluate Online Learning in the Context of University Training?

Fatima-ezzahra Mrisse<sup>1</sup>, Nadia Chafiq<sup>2</sup>, Mohammed Talbi<sup>3</sup>, Kamal Moundy<sup>4</sup>

Laboratory in Sciences Information and Education Technology (LASTIE)-Faculty of Sciences Ben M'Sick, Hassan II University, Casablanca, Morocco<sup>1,2,4</sup>

Observatory of Research in Didactics and University Pedagogy (ORDIPU)-Faculty of Sciences Ben M'Sick, Hassan II University, Casablanca, Morocco<sup>3</sup>

**Abstract**—This research presents online assessment in a digital device in the context of university training, aimed at improving their practices with emerging technologies based on an experiment with students from Hassan II University. Or, online assessment is a systematic process that helps measure the knowledge and skills of learners through multiple technological tools in a digital device. Indeed, digital devices intend to revolutionize higher education with the use of Information and Communication Technologies (ICT). Nevertheless, digital devices pose the problem of student identity verification during online assessment. In reality, automated online assessment systems are extremely vulnerable to cheating. So, our aim of this research is to explore, firstly, the types of online assessment that could be implemented in a digital device and secondly, how to verify the identity of the student during an online course on a digital device? The sample of our experiment consists of (N = 108) students from the Hassan II University of Casablanca, divided into two classes of the ITEF and MIMPA Masters and, our study was based on an online questionnaire for (N = 37) teachers at Hassan II University in Casablanca. The results obtained is to put into practice in digital devices diagnostic evaluations and formative evaluations using biometric methods for identity verification with a limited number. However, biometrics is inapplicable in summative assessments due to the problem of massiveness and hindrances in the online exam. For this reason, measures must be put in place to promote the smooth running of the online assessment.

**Keywords**—Online assessment; digital device; Information and Communication Technologies (ICT); biometrics; identify digital; student

### I. INTRODUCTION

Since March 2020 and due to the measures taken by the Moroccan authorities to combat the spread of COVID-19 such as the suspension of all activities involving a gathering of people, education in Morocco has undergone a remarkable evolution with the replacement of face-to-face courses by distance education. However, distance education allows students and all those who wish to train online and interact with others, particularly in the health crisis that has affected our country [1]. As well as classroom exams have been replaced by online assessments.

First of all, online assessment is a crucial step in distance learning in order to measure the degree of student acquisition through different assessment methods and in different fields (education, economics, engineering ...). This section aims to describe the theories from the literature review related to online assessment. Indeed, online evaluation should not simply invent new technologies that recycle our inefficient practices [2]. For example, the issue of assessing student learning in an online course has not been fully addressed [3]. And that assessment is at the heart of the teaching-learning process. What is assessed defines what is taught and how it is learned. The assessment process, in turn, shapes practice and affects the learner's view of the value of engaging in learning [4] and that online assessment remains an emerging new practice for most educators and trainers [5]. Furthermore, in professional settings and training institutions, assessment practices are numerous and diverse. Their most often stated purpose is to improve production processes and/or to certify the achievement of objectives [6]. Because of these observations, the authors underline the interest in online evaluation and the growth in its use, which is based on social, technological and economic factors. Assessment practices (diagnostic, formative, summative), whether face-to-face or online, can therefore be a mobilizing lever in an initial or continuing university education institution. These practices aim to achieve learning objectives through various authentic or collaborative activities in order to improve the development of students' skills [7].

Although Morocco is returning to its normal state thanks to the gradual decrease of the COVID-19 pandemic, distance learning retains a particular importance as it meets the growing and diversified needs of Generation Z learners, such as developing their skills and that they prefer to learn through digital resources and create their own personal learning environment [8,9]. It is now seen as a solution to ensure pedagogical continuity with learners in particular, those who cannot always be present. Hassan II University of Casablanca has developed several digital services grouped together at the Digital Workspace <https://ent.univh2c.ma>. This has allowed all the actors to have a simplified access to many institutional tools: electronic mail, e-portfolio, software, electronic books and magazines, e-learning platform. The digital resources of

the UH2C allow its students to regularly follow their courses in direct broadcast mode or in deferred access on the Moodle platform. The teachers of the Hassan II University of Casablanca use several digital devices in their teaching practice. Through these devices, teachers will be able to make their courses available online. However, it is difficult for the teacher to really practice summative evaluation online because on the day of the online exam, the learner can use any excuse to justify his absence or non-participation in the course.

According to previous research, students have been very satisfied with taking courses on online learning platforms and assessment methods [10]. Moreover, online and peer assessment plays a key factor in students' engagement during their learning and leads to learning performance [11]. The development of online assessment based on a standardized model helps to improve or enhance students' online learning [12]. Furthermore, the importance of integrating online assessment activities such as online knowledge surveys, online closed or open-ended questions, and peer review [13]. However, assessment activities should measure the degree to which students acquire learning objects with the same online opportunities and conditions [14].

It is for this reason that we are going to carry out this experiment, at the level of the Hassan II University of Casablanca, which will focus on formative evaluation practices. In this regard, the purpose of this research is to present the online assessment in a digital device in the context of university training, aiming to improve their practices with emerging technologies based on an experience with students of Hassan II University. In this context, we posed the following questions:

- What are the types of online assessment that could be implemented as part of a digital system?
- How to verify the identity of the student in the context of online assessments?
- What measures should be put in place to promote online assessment?

## II. THEORETICAL FRAMEWORK

This part aims to describe the theories resulting from the bibliographical review in relation to the types of online evaluation and the phenomena of cheating that could be seen during online evaluation.

### A. Online Assessment Practices

Generally, there are three types of evaluation, which can have different purposes. Most often, we distinguish:

1) *Diagnostic assessment or pre-test*: done before or at the start of a course or program to determine the prerequisites or knowledge required. A distinction is sometimes made between diagnostic and prognostic evaluation. The first concerns first of all the learner, allowing him to adapt his path or to take remedial measures and, the second serves rather the evaluator aims to produce information which makes it possible to orient or train or to adjust the training to his profile [15]. It makes it

possible to determine the profile of the learner in order to guide the course that adapts to his knowledge path.

2) *Formative assessment*: done during a course or program to support or enhance learning and foster learner motivation based on observed needs. It insists on the value of consolidation and feedback on error in order to support learning [16]. Its purpose is to check whether the learner is progressing and approaching the chosen educational objective of the established program. It promotes integrated motivation and encourages the learner to adopt a more effective approach to learning [17]. [18] Describe five functions of formative assessment: explaining goals, tasks and assessment, seeking evidence of understanding, seeking evidence of progress of the task, feedback anticipating future steps and empowerment of learners. However, the feedback therefore plays a central role. It is often more continuous or repeated at regular intervals during training.

These functions can be put to good use in the current situation of pedagogical changeover for the benefit of student learning.

3) *Summative assessment*: done after a course or program or during exams. This terminal practice aims to categorize, certify and validate practices, behaviors or knowledge [19], and to report. Its purpose is to say if such a learner is worthy of such a grade or if he can access the higher class. Consequently, it makes it possible to provide a balance sheet and to allow a decision (does he access the upper class or not?)

### B. Cheating

Cheating or copying<sup>1</sup> would therefore be associated with unauthorized mutual assistance, particularly with cheating in real time, often during exams. [7] For (CCA<sup>2</sup>, 2010), in the context of online assessment, for example:

Swap answers during an exam:

- By computer (instant messaging, email, etc.).
- By mobile phone.
- By text messaging.

And, according to [20] it gives the example of a student who “connects with the same username and the same password as another student to steal information or work” or, more simply, the one who obtains from a colleague the questions of a previous examination; “to consult non-admissible notes during an exam”. For example, his or her own notes or the web, when it is not allowed to:

- Submit work done in whole or in part by another person.
- Submit the same work, which may be an exam, for various courses without having obtained permission, where the work may be an exam.

<sup>1</sup> Defined by the great terminology dictionary of the Office of the French language as: “Fraud of a pupil who copies, during an examination or a test, the duty of a comrade, a book or course notes”.

<sup>2</sup> The Canadian Council on Learning.



However, cheating mostly targets contexts where external sources are not allowed. Additionally, plagiarism or copying generally applies in the context of a work where the use of other sources is permitted, but on condition that credit should be clearly attributed to the original author. What are the cheating detection tools that can be implemented in a digital device?

### III. METHODOLOGY

#### A. Measuring Instrument

To answer the above questions, we opted for a quantitative methodology through a questionnaire. We drafted a questionnaire for teachers and then administered it to 37 teachers at Hassan II University in order to enrich our problematic.

#### B. Methods of Experimentation

Therefore, we searched for suitable tools and methods to address the aforementioned issue. So, we have chosen to integrate the biometric method which makes it possible to specify the automatic identification of a student during online evaluations. The latter measures the uniqueness of an individual from the unchangeable parts of his body. This method is based on two methods: Behavioral modalities and, morphological modalities. In addition, we decided to work with a morphological modality through the recognition of the student's face and the other behavioral modality through the recognition of the student's voice. Then, we set up a system with emerging technologies. Next, we opted for the latest version of the moodle platform because it is more suitable for the implementation of a real digital device than the Open edx platform. In addition, we installed the moodle platform online with the xamp server, by integrating two plugins: one to block the search browser window during the online assessment. And the other which is based on the camera and the microphone to ensure the online identification of the students.

#### C. Echontillon

To answer the research questions above, we carried out an experiment with the students of the Master class of Engineering and Technologies for Education and the Master class Instrumentation and Physico-Chemical Methods of Analysis in second semester of the quality module: Management and tools at Hassan II University. The number of our sample was (N=108) and our experiment on the digital device [www.digitaloftkill.net](http://www.digitaloftkill.net) (Fig. 1) which was in the form of a questionnaire of 20 multiple-choices questions. Each question with four answer choices in a period of 15 min.

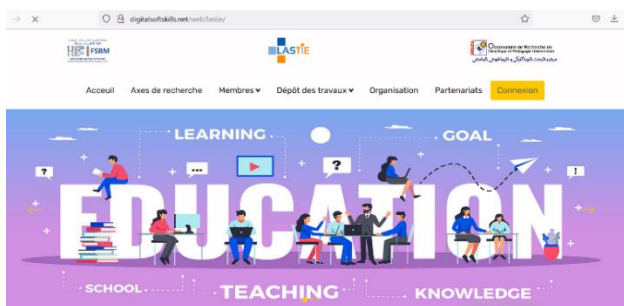


Fig. 1. Overview of the platform [www.digitaloftkill.net](http://www.digitaloftkill.net).

We have created an account for each student that contains a username and password in order to take the online assessment. Each student must follow the instructions below before starting the evaluation, respecting the period of time determined and the number of possible attempts (Table I).

TABLE I. TABLE OF INSTRUCTIONS

Instructions
1. Install Safe exam browser version 3.1.1 on your desktop.
2. Inform people around you not to enter the place where you are doing the MCQ.
3. Download the QCM SEB file.
4. Click on the meeting link.
5. Activate your camera and sound.
6. Click on the test file (for example: mock test).
7. Click on no.
8. Enter your following username and password: Username: student2 Password: Student2 @ 21
9. Click on continue the attempt (to start the test).
10. Answer the exam questions.
11. Click on end test.
12. Click on send all and finish.
13. Click on quit Safe exam browser.

### IV. RESULTS ANALYSIS

#### A. Results of the Questionnaire

To answer the questions above, we wrote a questionnaire for the teachers then, we administered these questionnaires for 37 teachers of the university Hassan II in order to enrich our problem. Based on this teacher questionnaire, we present some of the following results:

In the situation of the COVID-19 pandemic, we noticed that there are 60% of teachers preferred online courses; on the other hand, there are 40% of teachers preferred face-to-face courses (Fig. 2).

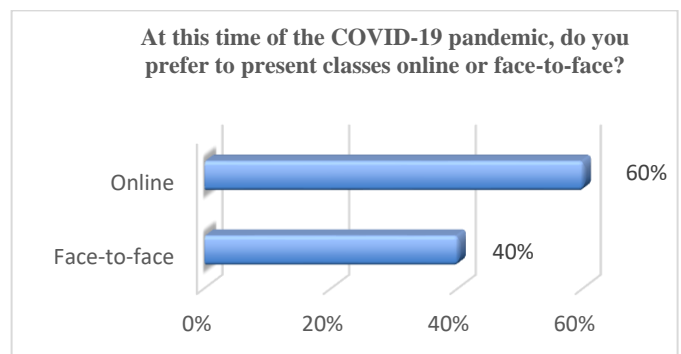


Fig. 2. The choice of online and face-to-face teaching.

From this diagram, we found that 38% of teachers prefer online assessment and 62% of teachers prefer face-to-face assessment (Fig. 3).

We noticed that there are 65% of teachers do not trust students during online assessment and 35% of teachers trust students (Fig. 4). Hence the question: What measures should be put in place to promote online assessment?

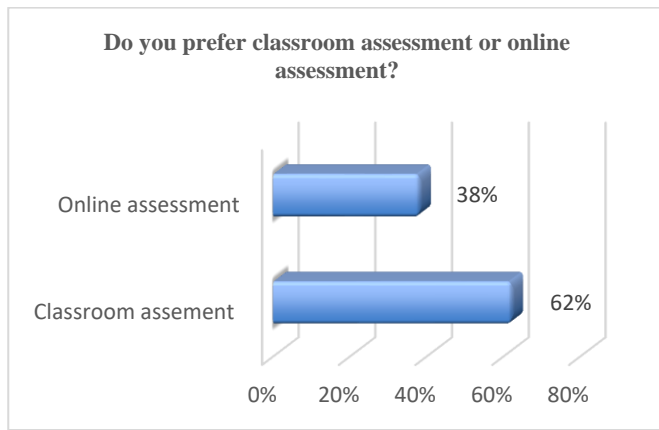


Fig. 3. The choice between in-class and online assessment.

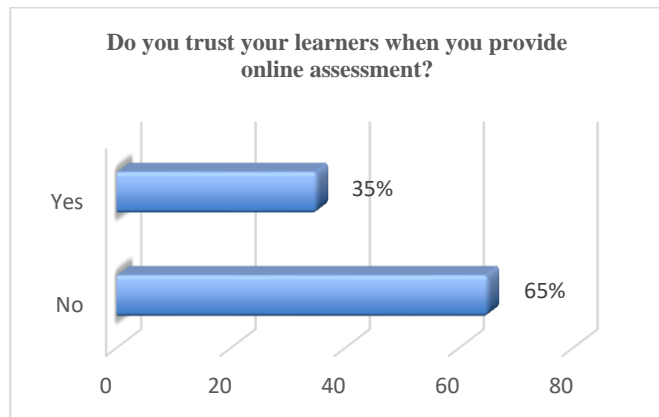


Fig. 4. Teacher confidence in online assessment.

### B. Experimental Results

We present below the results of this experiment: these are general data relating to the gender questionnaire, average scores, number of attempts and the constraints (or obstacles) that hinder online assessment practices, in order to determine the measures to promote the establishment of an online assessment.

Regarding the percentage of students, we have 60% of female respondents and 40% of student respondents, which shows that female respondents are more numerous than male one (Table II).

Regarding the average pass mark, we have 14.57 for female students and 13.63 for male students and 14.16 for the general average of the two classes. So, we can deduce that the majority of students have good grades and have not found any difficulties in the MCQ. Although it is necessary to take into consideration the students who failed and those who could not pass the MCQ (Table III).

TABLE II. CHARACTERISTICS OF THE SAMPLE BY GENDER

Gender	Percentage	Number of participants
Female	60 %	64
Male	40 %	44
Total	100%	108

TABLE III. THE AVERAGE SCORES ON A PASSAGE TEST

The average score	
Female	14.57
Male	13.63
General average	14.16

Regarding the success rate of the students, we have 86.11% of the students who passed and, 5.56% of the students failed in the assessment as well as, 8.33% of the students could not pass the assessment. So, we can deduce that most students have succeeded in this online assessment, but we must try to understand the difficulties (or the obstacles) encountered by other students in online assessment practices (Table IV).

TABLE IV. THE STUDENT SUCCESS RATE

Success rate	Percentage	Number of participants
Passed	86,11%	62
Failed	5,56%	6
Did not pass the assessment	8,33%	9
Total	100%	108

Among the 99 students who passed the MCQ, we find that 48.49% of the students who answered on the first attempt while we have 51.51% of the students who answered after the 2nd attempt because they encountered obstacles during the 1st attempt (Table V).

TABLE V. THE NUMBER OF STUDENT RESPONSES ACCORDING TO ATTEMPTS

Responses by attempts	Percentage	Number of participants
Student responses after the 1st attempt	44,44%	48
Student responses after the 2nd attempt	47,22%	51
No answer	8,33%	8
Total	100%	108

### C. Obstacles Encountered

According to this experiment, we find that 5.56% of students who did not follow the instructions requested. We also have 2.78% of the students who encountered hardware problems; they didn't have cameras in their computers. We also note that 19.44% of the students found problems installing the Safe exam browser because this software requires a more recent operating system whereas the students had an older version on their computers (windows 7, windows XP, etc.). We also have 5.56% of students who found constraints in the quality of the internet connection as the quality of the internet network is low or limited because, the connection which is consumed during the evaluation. Then, we have 13.89% of the students received error messages indicating that the server becomes slow when it is loaded because of a large number of users. On the other hand, we see that 44.44% of the students did not find any problems during the online evaluation on the first attempt (Table VI).

TABLE VI. THE CONSTRAINTS EXPERIENCED

Constraints	Percentage	Number of participants
Problem of following requested instructions	5,56%	6
Material problem: Camera	2,78%	3
Connection quality problem	5,56%	6
Problem installation of the browser secure exam due of the update operating system (Windows 7 and Windows XP)	19,44%	21
Server load problem	19,44%	21
Didn't find any problem	44,44%	48
Total	100%	108

## V. DISCUSSION

For the question “What types of online assessment could be implemented in the context of a digital device?”, we always advise teachers to implement diagnostic assessments and formative assessments on digital devices. However, it is not recommended to set up summative evaluations there because of the obstacles that can be encountered during online exams such as: (the loss of light, the poor quality of the Internet network, the overload of the server quality or the error messages that occur following the overload of the server, malfunction of the hardware equipment for example: the camera, the microphone, etc.).

For the question, "How to verify the identity of the learner in the context of online assessments?", we can verify the identity of the student with the biometric methods in front of the camera and by voice recognition with the microphone within the framework of diagnostic and/or formative assessments with a limited number of students. However, biometrics is inapplicable in summative assessments due to the problem of massiveness and the constraints that hinder online assessment. Regarding the measures to be put in place to promote the smooth running of the online assessment, we can mention:

### D. Verification of Student Identity during Online Assessment on a Digital Device

Among the 99 students who responded at to the 1st attempt or the 2nd attempt, the identity of 75.76% of the students was verified whereas we could not verify the identity of 24.24% of them (Table VII).

Among the 24.24% of students whose identity could not be verified, 21.21% of them were not verified by the Safe exam browser tool and 3.03% of other has not been verified by camera and microphone (Table VII).

Regarding the verification of the identity of the student by face and voice recognition, we verified with 96.97% students who have the camera and the microphone through the Google meet tool. But, we could not verify with 3.03% who do not have the camera and the microphone (Table VIII).

For the phenomenon of electronic cheating, we found the SEB tool (which allows to block the search browser window) to be feasible with 78.79% of students who were able to install SEB but, is not feasible with 21.21% students (Table IX).

- Design very clear instructions to guide students in preference, to be given before the day of the assessment or to take a practical test.
- Ask the learners to prepare the appropriate equipment for the smooth running of the assessment (a good camera, a good microphone).
- Require students to have good internet quality (a speed that allows the assessment to be taken without internet interruption) or to provide a good internet network to all students at the time of the online assessment.
- Ask students to update their operating system at the beginning of the module.
- For teachers, it is recommended to use online servers and avoid using local servers.
- Declare the time and period of the online assessment and, make more than one attempt to overcome all obstacles during the online assessment. For teachers, it is recommended to use online servers to avoid error messages from overloading local servers.

TABLE VII. VERIFICATION USING TECHNOLOGICAL TOOLS

Verification by technologiques tools	Percentage	Number of participants
Couldn't verify by biometric modalities (camera, microphone)	3,03%	1
Couldn't verify by Safe exam browser tool	21,21%	7
Total	24,24%	8

TABLE VIII. VERIFICATION BY FACE AND VOICE RECOGNITION

Verification by face and voice recognition	Percentage	Number of participants
Don't have a camera and microphone	3,03%	1
Have camera and microphone	96,97%	32
Total	100%	33

TABLE IX. ASSESSMENT BY THE SAFE EXAM BROWSER TOOL

Assessment by the Safe Exam Browser tool	Percentage	Number of participants
Couldn't install SEB	21,21%	7
Were able to install SEB	78,79%	26
Total	100%	33

## VI. CONCLUSION AND FUTURE WORKS

This research aims to present the online assessment in a digital device in the context of university training, aiming to improve their practices with emerging technologies based on an experience with students of Hassan II University. Our discussion to the results that we recommend teachers to

implement diagnostic assessments and formative assessments on digital devices and to avoid implementing summative assessments on them. Thus, we can verify the identity of the student with biometric methods in front of the camera and by voice recognition with the microphone in diagnostic and/or formative assessments with a limited number of students. But biometrics is not applicable in summative evaluations.

As a follow-up to this study, we plan to analyze students' and teachers' attitudes about the use of biometrics on learning platforms during the learning of a given module and with a focus on passing assessments.

#### ACKNOWLEDGMENT

The corresponding author would like to thank the Faculty of Sciences Ben M'sick, Hassan II University, for their support in experimenting with this research, as well as all the teachers and the students who participated in this study. This research received no external funding. The authors would like to thank the editor and reviewers for their insightful comments on the first draft of the article.

#### REFERENCES

- [1] T. Karsenti, B. Poellhuber, N. Roy, and S. Parent, "Le numérique et l'enseignement au temps de la COVID-19 : entre défis et perspectives – Partie I," *Rev. int. technol. pédagog. univ.*, vol. 17, no. 2, pp. 1–4, 2020.
- [2] M. Ripley, "E-assessment question 2004," Presentation at Delivering E-assessment - a Fair Deal for Learners, QCA, April 20, cited in Ridgway and McCusker, 2004.
- [3] M. Robles and S. Braathen, "The practices and challenges of online assessment," *Francophone Distance Education Network of Canada*, March 2011, p 6.
- [4] JISC, "Effective Practice with e-Assessment," *The Practices and Challenges of Online Assessment*, *Francophone Distance Education Network of Canada*, March 2011, p 6.
- [5] H. Wang, "In his review of Howell and Hricko's (2006)" book in *The Quarterly Review of Distance Education*, volume 8, no. 4, 2007.
- [6] L. Paquay, C. Van Nieuwenhoven, P. Wouters, "Evaluation, lever of professional development?," *Tensions, systems, new perspectives: Pedagogies in development*. August 2010.
- [7] L. Audet, *Les pratiques et défis de l'évaluation en ligne*. Canada: Réseau d'enseignement francophone à distance du Canada, 2011.
- [8] D. Tonkery, *E-books come of age with their readers*. *Res. Inf.* 24, 26–27. 2006.
- [9] Rowlands et al., "The Google generation: the information behaviour of the researcher of the future," *Aslib Proc.*, vol. 60, no. 4, pp. 290–310, 2008.
- [10] N. Almusharraf and S. Khahro, "Students satisfaction with online learning experiences during the COVID-19 pandemic," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 21, p. 246, 2020.
- [11] M. Ma and C. Luo, "The effect of student and peer assessment engagement on learning performance in online open courses," *Int. J. Emerg. Technol. Learn.*, vol. 17, no. 10, pp. 145–158, 2022.
- [12] D. Baneres, X. Baró, A.-E. Guerrero-Roldán, and M. E. Rodriguez, "Adaptive e-Assessment System: A General Approach," *Int. J. Emerg. Technol. Learn.*, vol. 11, no. 07, p. 16, 2016.
- [13] B. Bahati, U. Fors, P. Hansen, J. Nouri, and E. Mukama, "Measuring learner satisfaction with formative e-assessment strategies," *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 07, p. 61, 2019.
- [14] J. Khlaisang and P. Koraneekij, "Open Online Assessment Management System platform and instrument to enhance the information, media, and ICT literacy skills of 21st century learners," *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 07, p. 111, 2019.
- [15] C. Hadji, "Evaluate, rules of the game," E.S.F. editor, Paris, 6th edition, 2000.
- [16] S. Dehaene, "To learn! The talents of the brain," the challenge of the machines. Paris: Odile Jacob, 2008.
- [17] J. H. Mc Millan and J. Hearn, "Chronological analysis of the journaled traces of a study for self-directed learning," 2008.
- [18] M. Grangeat, and C. Lepateur, "Roles of teacher feedback on the self-regulation of learning," *Assess, International Journal of Research in Education and Training*, 5(2), 5-28, p. 26, 2019.
- [19] J. Ardoino and G. Berger, "The Practices and Challenges of Online Assessment," *Francophone Distance Education Network of Canada*, March 2011, p 10, 1986.
- [20] N. Perreault, "The Practices and Challenges of Online Assessment," *Francophone Distance Education Network of Canada*, March 2011, p 53, 2007.

#### AUTHORS' PROFILE

**Fatima ezzahra Mrisse:** Doctor-researcher at Hassan II University of Casablanca in Morocco, Laboratory in Sciences Information, and Education Technology (LASTIE), Faculty of Sciences Ben M'Sick, Observatory of Research in Didactics and University Pedagogy (ORDIPU), Casablanca, Morocco.

**Nadia Chafiq:** Professor at Hassan II University of Casablanca in Morocco, Laboratory in Sciences Information, and Education Technology (LASTIE), Faculty of Sciences Ben M'Sick, Observatory of Research in Didactics and University Pedagogy (ORDIPU), Casablanca, Morocco.

**Mohammed Talbi:** Professor at the Hassan II University of Casablanca in Morocco, Laboratory in Sciences Information, and Education Technology (LASTIE), Faculty of Sciences Ben M'Sick, Observatory of Research in Didactics and University Pedagogy (ORDIPU), Casablanca, Morocco.

**Kamal Moundy:** PhD student at the Laboratory of Sciences and Technologies of Information and Education (LASTIE), Faculty of Sciences Ben M'Sick, Hassan II University, Casablanca, Morocco.

# Comparative and Evaluation of Anomaly Recognition by Employing Statistic Techniques on Humanoid Robot

Nuratiqa Natrah Mansor<sup>1</sup>, Muhammad Herman Jamaluddin<sup>2</sup>, Ahmad Zaki Shukor<sup>3</sup>, Muhammad Sufyan Basri<sup>4</sup>  
Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Durian Tunggal, Melaka, Malaysia<sup>1, 2, 3</sup>  
MAP2U Sdn Bhd, PT9951, Jalan BBN 1/3K, Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia<sup>4</sup>

**Abstract**—This paper presents the study to differentiate between normal and anomaly conditions detected by humanoid robots using comparative statistics. The study has been conducted in robotic software as a platform to examine the scenario and evaluate between the anomalies and normal behaviour in different conditions. This study employed a machine vision technique to run an image segmentation process and carry out semi-supervised object training within a controlled environment. The robot is trained by differentiating the measurement size of the target object, its location, and the object's visibility within three different frames. The effect is measured by extracting the positive predictive value (PPV) value, mean and standard deviation value from the captured image using statistical techniques in machine vision. The results showed that the mean value decreased by around 50% from the normal scenario when an anomaly occurred. Aside from that, the standard deviation values were more than twofold compared to the common scenario, especially after the object's size grew. In contrast, the deviation value is remarkably small when the target is situated in the middle of adjacent frames, compared to the value when the entire shape is positioned in the frame. Simultaneously, the mean values from the processed image produced a minor difference.

**Keywords**—Anomaly detection; humanoid robot; vision system; statistical analysis; robot recognition

## I. INTRODUCTION

Over time, the application and comprehension of machine vision and robotics research have become increasingly valuable and helpful, especially for humans. Among the numerous studies, one that has piqued researchers' interest is applying picture segmentation approaches for spotting objects and people in various settings and orientations. Hence, it is scientifically referred to as anomaly detection. Anomaly is also defined as a low likelihood or probability of occurrence in a given setting. To put it simply, anomalies occur when the item's state deviates from the standard or customary order [1,2]. Certain anomaly detection studies generally implemented computational intelligence knowledge, including fuzzy logic-based outlier identification and neural networks.

Moreover, this detection algorithm employs machine learning and statistical detection methods, as the subject's previous behaviour or trends need to be learned and interpreted [3,4]. It may also be utilized for practical purposes in product quality control, enhancing national security, and

military purposes. Anomaly detection might also be used for mobile robots and unmanned aerial vehicles (UAV). These robots can detect anomalous behaviour and sceneries in various environments. Among them, the humanoid robot is a type of mobile robot. The robot can be programmed to use picture segmentation and processing techniques to determine the output pixel value and identify anomalous behaviour.

In anomaly detection, there are several issues that need to be considered, such as objects' features and illumination circumstances [5], dense 3D shape models, image resolution, objects in different frames, and many other things [5-7]. Although some of the methodologies and procedures worked well under specific conditions or requirements, the accuracy for distinguishing anomalous behaviour has significant degradation due to different issues. Anomaly detection is perhaps one of the promising and fascinating fields to study.

Among the issues, the issue of recognizing static objects when located in different positions is addressed in this study. Apparently, the issue occurs when there is confusion regarding the identical frames and variable object sizes. Besides that, the issue also arises when an object is placed at the boundary or when the object is located in a different position or corner in the frame. The object may not appear identical to its predecessor as it holds the corresponding pixel value in every scenario. Mean values and deviation of pixels value also vary between photos. Hence, this research focuses on utilizing humanoid robots by integrating machine vision studies to detect any irregularities in their view and classify them using statistical techniques. This project was created using the CoppeliaSim and Python (x,y) programming to demonstrate the method and simulation of how the work functions.

The next section will delve into the broad idea of machine vision with the theories of robotics vision and humanoid vision as part of the component with the block diagram of necessary processes in the system, including techniques for learning in anomaly conditions. Then in the methodology part, the procedures and methodologies performed on each experiment are presented in detail. In the following section, all recorded data and information observed from every single experiment are tallied and illustrated in graph form. Next, Section V will discuss and evaluate the outcome after the simulation of each experiment to support the hypothesis and proposed method. Meanwhile, Section VI summarizes the key

findings and conveys realistic recommendations for improvements in future works.

## II. LITERATURE REVIEW

### A. Machine Vision

Visual exploration in unknown settings is a common and crucial job for intelligent automation in the present day. One of the most compelling aims for machine vision applications is for robots to be able to search for identifiable things in an unknown environment; even though some robots have a deficiency in visual skills because of unable to compete with human ability, particularly in brains and eyesight [8]. Regardless, it may still be explored and improved regularly with various concepts and theories, especially with the continuation of innovative technology.

Machine Vision is a “simple” processing system that captures visual information by processing, analyzing and measuring the image attributes using a combination of hardware and software. This is used to extract the information utilized in decision-making in order to generate the required result. The application of vision has been developed into a software platform to make people unnecessary spending an extended time developing algorithms for image processing [9].

Furthermore, the platform is user-friendly and allows for a simple extraction and modification of image information. Some of the examples are HALCON from MVTec Software, IMAQ from National Instruments, Visilog from Norpix, and PatMAX from Cognex Corporation [10]. It may sometimes surpass humans in specialized tasks. As shown in Fig. 1, the practical strategy for recognizing objects in vision systems mainly uses local characteristics as a primary tool to match scene material to models of recognized items.

### B. Robotics Vision

Robots are growing increasingly capable in line with the decade’s developments because they can execute a growing number of activities that would be too risky or difficult for humans. Flexibility has been required as job complexity has risen. The need for adaptability has grown in tandem with the complexity of modern jobs. The creation of humanoid robots gives way to the establishment of societies of robots that collaborate with humans as humanoid robot technology advances [9,10]. The evolution of robots goes from single and tiny robots to the establishment of humanoid and gigantic robots that can work together with humans.

The application of vision is becoming a critical element in robotics and machine automation. Hence, the studies have been implemented in robotics advancement, known as robotic vision. In standard function, robot vision is utilized for part identification and navigation. In advanced function, the vision enables to lead and guide the robots or other automated machines to recognize, think and perform the task, resembling a human [11,12]. By controlling the special machine vision software, it can make the part identification and replace or complement manual inspections, navigation and measurements with digital cameras and imaging processes. Then the controller will automate and coordinate a necessary

process to perform the task. The general process can be seen in Fig. 2.

### C. Humanoid Vision

Ideally, the vision has been applied to many types of robotics and machines, not being left behind in humanoid robots. With the intelligent vision system used in humanoids, the robot can interpret images in terms of colour, size, shape, and position of the object on the scene. In addition, both sights of a humanoid robot can recognize various objects with different traits. The image will undergo pre-processing and image segmentation stage before continuing with object matching and tracking [13, 14]. An intelligent system is also able to perform a high-level recognition task. Some of the essential elements in the automated analysis and extraction of the required information are shown in Fig. 3.

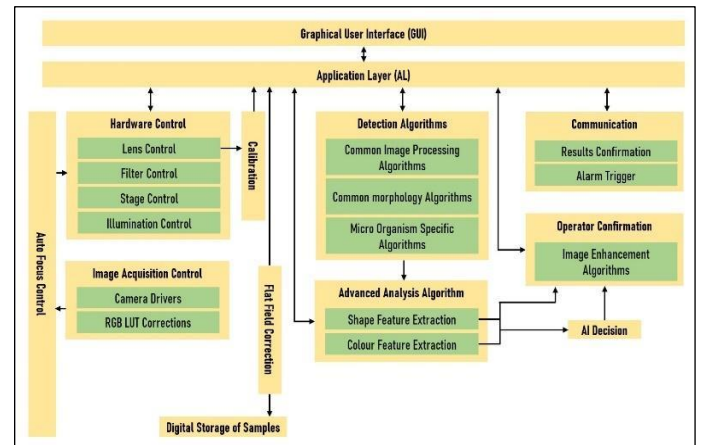


Fig. 1. Machine vision software architecture.

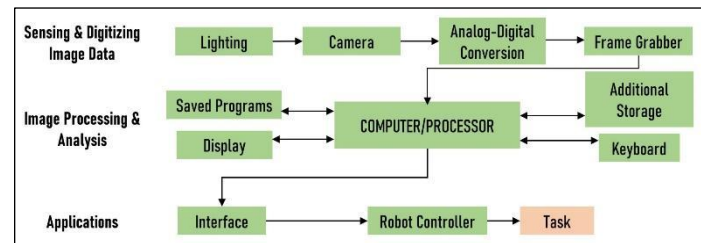


Fig. 2. Machine vision process with three important components.

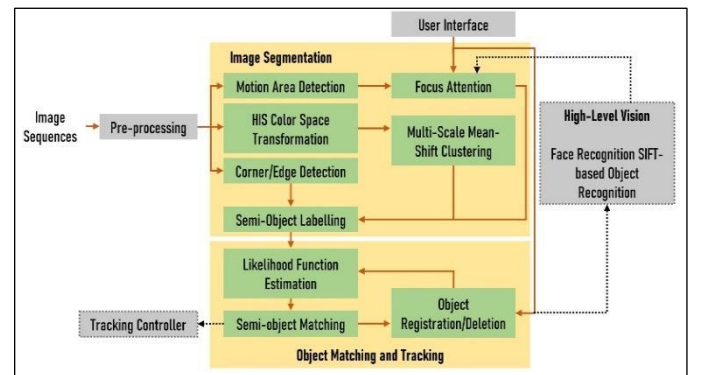


Fig. 3. Structure of the robotic vision system.

#### D. Methodology Used

Anomaly detection can occur in different surroundings and circumstances, depending on what abnormal behaviour are set. Therefore, anomaly detection methods are classified according to their input type, output type, learning method and the type of target anomaly. There are three types of learning methods in anomaly detection, which are supervised, semi-supervised and unsupervised [9,10].

For supervised learning, data sets labelled with normal or anomalous are used for training the model. While for semi-supervised, the model is trained with only normal data. The data points that do not match normal behaviour are considered anomalous during the detection. Differ for unsupervised learning case, no training data required. The input data is expected to contain a high ratio of normal to anomalous points.

Alternatively, the model will consider frequent anomalous behaviour as a common condition [15]. As mentioned earlier, the main objective of this research is to detect an anomaly behavior of non-motion objects via a humanoid robot. Besides that, the anomaly behavior in different frames is also observed using the single camera attached to the robot. The experiment is handled in a roomy space in which all the variables are controlled, fixed and not moving.

Thus, the method highlighted earlier is reasonable and much more apposite to be carried out in this project for anomaly behaviour as it can detect objects even though in a different position of a frame. First, this method is tested on the same base background and light intensity with different object sizes. After that, the outcome from each experiment and observation is turned into the percentage of accuracy, precision, and limitations. Finally, the data is discussed in the analysis and result section.

### III. METHODOLOGY

Generally, this study aims to let the humanoid robot learn to detect and identify objects in anomaly behavior. In order to observe and evaluate the performance of object identification and anomalies cases, a series of experiments is conducted to prove the object's behaviour in different cases. Primarily, specific aspects and parameters need to be determined and controlled to maintain the accuracy of the experiments. As an example, lighting and frame background. Originally, the methodology to conduct the experiment was broken down into two stages. The first one is the system design phase which discusses about the software and type of simulation used to run the experiment. The other stage explains how the methodology is implemented, where the details on how experiments are handled to get reliable results.

#### A. System Designation

The experimental platform to work for this project is Python (x,y) for Windows and CoppeliaSim software. This study is developed and runs through the simulation process by these two softwares. The coding code for the programming is developed in Python. At the same time, the environment scene and robot used for the experiment are built in a robotic simulator, CoppeliaSim. These two softwares are linked to

each other and act as terminals to show the overall simulation and experiment process.

In overall, this project undergoes four different types of experiments. The experiments are semi-structured scenes, calculating the mean and the standard deviation values based on different sizes of the object, determining the location and position of the object from the robot and lastly, collecting the mean and standard deviation values based on the different positions of an object in one frame.

#### B. Stereo segmentation

Interest points that are matching in the left picture are divided into groups based on their location and discrepancy. Then, a greedy clustering approach is applied, starting with a single interest point. After that, if the x-position, y-position, and disparity are within the 5-pixel threshold of any neighbouring cluster member, it is supplemented with new ones. The segmentation usually occurs at a comparable depth but in a different spatial position or at a nearby spatial location but at a different depth. Fig. 4(a) and 4(b) showed the pattern of the process.

To enhance the segmentation step, feature locations are divided into on-object and outside-object groups (white vs other colours). According to the above image in Fig. 4(a) and 4(b), every bit of outside-object locations and any image pixels outside the bounding region are considered outside-object pixels. This region is emphasized in black colour after the segmentation process. Following that, the location for the on-object is highlighted in grey colour. To indicate that the particular shade is an anomaly, an image's pixel value is generated in histogram form so that it is easier to compare between the normal and abnormal or determine the different cases for every anomaly.

#### C. Anomaly Criterion

While learning the object samples, two sorts of anomalies can be identified; For starters, the event patterns that never have been seen before and employed as variable for classification. The flow is shown in Fig. 5 below. The anomaly scenario is defined as samples that deviate significantly (in terms of distance) from other clusters in the model [16]. Secondly, the model contains exemplars of other sorts of anomalies. It does, however, have a minimum frequency this time. As a result, it might also be characterized as a false positive. Both forms of abnormalities may be discovered by taking distance and frequency into account to classify.

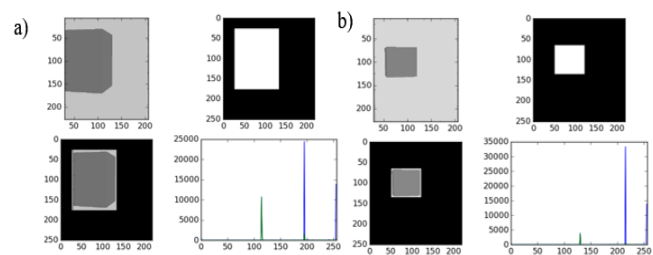


Fig. 4. Image segmentation, where the feature for matching exists, is clustered and plotted into the histogram.

According to equation (1), an intra-class distance  $Dist(C_i)$  in each cluster is defined as the greatest distance between any cluster member and the cluster centre.

$$Dist(C_i) = jmax\|f(e_j^{C_i}) - f(C_i)\|_2 \quad (1)$$

Thus, as long as the abnormal grid event is detected, the frame triggered the condition as abnormal, either in significant or small deviations [16]. The deviation is calculated by using a simple formula. In this study, the deviation is extracted using the plug-in Open CV and Numpy Software in Python (x,y).

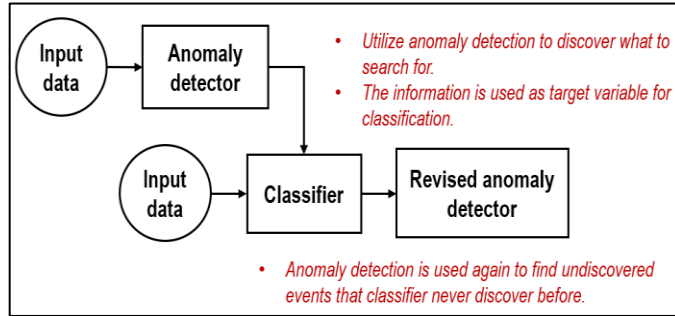


Fig. 5. Example of flow process of anomaly behavior for classification.

#### D. Research Procedure

1) *Experiment A: Semi-supervised Scene:* This experiment investigates the object's behaviour and the pixel number when it is located at a different position in the same black background frame. The object is fixed in a triangular shape and red colour. The output results were positive predictive value (PPV), which is technically the pixel data extracted from the image captured. In addition, intra class-distance is calculated manually to find the percentage of anomaly that can happen. Finally, the value is compared with the smallest distance value to identify true-positive and false-positive anomalies.

2) *Experiment B: influence on the object size:* This experiment measures the performance and effect of the variant size of an object on the system. The object that has been used is a cuboid with the same colour as the previous experiment. The frame background is fixed to a grey colour to differentiate it from black. Black is the default colour for the surroundings, as seen on the vision sensor recorded in the CoppeliaSim. There are three 50 cm-width and 80 cm-high walls labelled as Frame 1, Frame 2 and Frame 3. Frame 1 is set at the origin at  $180^\circ$ . All frames have the same surface, colour and size.

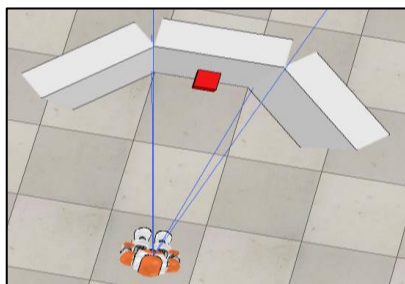


Fig. 6. The top view of the experimental setup is placed ahead of the NAO robot.

Meanwhile, Frame 2 is located at  $-40^\circ$  and deviates from the horizontal axis. Meanwhile, Frame 3 has deviated as much as  $40^\circ$  from the horizontal axis. The setup can be seen in Fig. 6 above. The parameter for the object's size is set from 0.4 cm x 0.4 cm up to 45 cm x 45 cm. This dimension is for Frame 1. However, for Frame 2 and Frame 3, the object size added an extra size of 50 cm x 50 cm. This is because both frames need a larger cuboid surface to cover up all areas of the wall frame.

3) *Experiment C: detection of object's location:* This experiment is conducted to locate the anomaly by calculating the distance and angle between the robot and the wall screen. The experiment scene setup is still the same, as shown in Fig. 6. In this experiment, the cuboid's maximum size was not identical to the cuboid in Experiment B. The maximum size of the cuboid for Frame 1 and Frame 3 is 20 cm x 20 cm, while for Frame 2 is 25 cm x 25 cm, depending on all cuboid sides that are possible to be fitted in a frame. Using the formula of Pythagoras' Theorem and Sine Rule, the location of the cuboid in each frame can be ascertained. The experiment is continued with the same procedure in each particular frame.

4) *Experiment D: impact on various placement:* From the result obtained in experiment B, only one size of a cuboid is selected as a fixed parameter in each frame. The cuboid size is picked based on the output mean and standard deviation difference between normal and abnormal conditions. For this time, the middle range of difference is selected. The cuboid is placed independently from 0 cm to 50 cm along the wall. Then, the effect of object position on the system is measured from the data attained in the mean and standard deviation of the pixel value. The experiment is continued with the same procedure in each particular frame. The difference between the normal and anomaly condition is observed based on the data outcomes.

## IV. RESULTS AND DISCUSSION

### A. Semi-Structured Scenes (Semi-Supervised)

For the early experiment, only one frame is used despite the different positions of the target. The reason for this is to observe the performance of the target object. In this case, a triangular object is placed on the screen. The black background colour is chosen to make the number of pixels easy to calculate, as this experiment has not yet applied the image segmentation process. The object has been used as the target and labelled by 'A'. Even though the object is in different positions on the frame, 'A' is successfully recognised and marked as normal behavior. The series of the captured image by the robot vision is shown in Fig. 7.

Normal scenes are classified whenever the object is still inside the frame range, not at the edge or outside the boundary. From the early hypothesis, a subject is marked as an anomaly when it is located at the frame's edge. The reason for this is that the frame's pixel value is counted differently than it is supposed to be. However, depending on the situation and output data, it can be categorized as a false positive or a true positive.



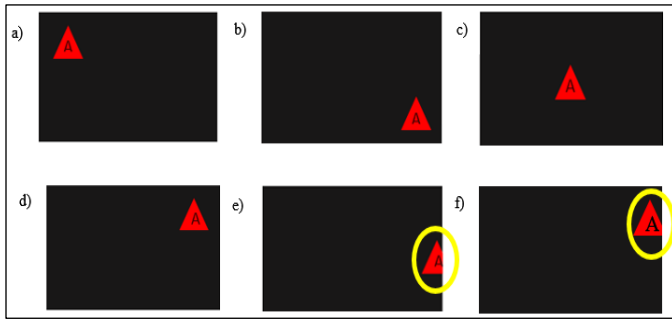


Fig. 7. A collection of object samples observed by the camera in a particular circumstance. Images (a)-(d) depict normal behaviour, whereas images (e)-(f) depict anomalous behaviour.

Precision or positive predictive value (PPV) is the ratio of the number of object pixels that are correctly recognized object pixels to the number of detected object pixels. When the ratio computation is near one, the performance is good. It is also known as the True Positive Rate (TPR), where the ratio of the number of object pixels correctly acknowledged to the number of object pixels in the ground truth. These quantitative analyses can be seen in Table I.

TABLE I. PPV VALUE OF OBJECT SAMPLES IN DIFFERENT POSITIONS

Sample	1st	2nd	3rd	4th	5th	Average PPV Value
a	0.95	0.96	0.97	0.96	0.96	0.96
b	0.94	0.95	0.94	0.94	0.94	0.94
c	0.98	0.99	0.98	0.99	0.98	0.98
d	0.95	0.94	0.95	0.95	0.95	0.95
e	0.67	0.67	0.67	0.67	0.67	0.67
f	0.87	0.88	0.89	0.89	0.89	0.89

$$PPV = \frac{\text{number of object pixels is correctly identified}}{\text{number of object pixels detected}} \quad (2)$$

By calculating the mean, variance and standard deviation using the formula follows:

1) Mean,  $\mu$

$$\mu = \frac{\sum x}{N} \quad (3)$$

$$\mu = \frac{0.96+0.94+0.98+0.95+0.67+0.89}{6} = 0.8983$$

2) Variance,  $\sigma^2$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (4)$$

$$\sigma^2 = \frac{1}{6} [(0.96 - 0.8983)^2 + (0.94 - 0.8983)^2 + (0.98 - 0.8983)^2 + (0.95 - 0.8983)^2 + (0.67 - 0.8983)^2 + (0.89 - 0.8983)^2] = 0.0112$$

3) Standard Deviation,  $\sigma$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (5)$$

$$\sigma = \sqrt{0.0112} = 0.106$$

4) Intra - class distance,  $Dist(C_i)$

$$Dist(C_i) = jmax \|f(e_j^{C_i}) - f(C_i)\|^2 \quad (6)$$

$$Dist(C_i) = \|0.67 - 0.945\|^2 = 0.0756 \approx 0.08 \quad (\text{for case } e)$$

$$Dist(C_i) = \|0.89 - 0.945\|^2 = 0.0030 \quad (\text{for case } f)$$

5) Smallest distance between  $f(e_x)$  and  $f(C_x)$

$$d(e_x, C_x) = 0.98 - 0.945 = 0.0035$$

The parameters that have been calculated above show that the value of standard deviation,  $\sigma \approx 0.1$  and intraclass distance is 0.08. Thus, it shows that the percentage of anomalies that can be occurred is very low. To prove the anomaly in Experiment 1, the calculated value from  $d(e_x, C_x)$  is less than  $Dist(C_x)$ , which shows that abnormality in Fig. 7(e) has happened. However, the value of  $Dist(C_x)$  and  $d(e_x, C_x)$  in Fig. 7(f) is much alike. Therefore from Fig. 7(e) and Fig. 7(f), two out of six cases are abnormal. Fig. 7(f) can be labelled as a false positive as the PPV value is very near to the PPV reading of ordinary cases. Meanwhile, the image in Fig. 7(e) is quite distant from the mean value and other ordinary cases. Thus, Fig. 7(e) can be labelled as true-positive.

Aside from that, the shape in Fig. 7(e) showed a bigger proportion surpassing the boundary of the frame compared to the smaller proportion of the object in Fig. 7(f). The percentage of the whole object in Fig. 8(f) is about 90%, but in Fig. 7(e) is almost 60%. Data that does not match normal behavior in semi-structured scenes is considered anomalous during the detection.

Based on experiment A that has been done, the pixel values are close to 1 despite the object's different position on a frame. The PPV value is significant when the object is at the edge of a frame. This value is categorised as an anomaly, whether false-positive or true-positive. False-positive is when the condition is an anomaly. Still, the output data is near the normal condition. At the same time, true-positive can be labelled when the condition is obviously an anomaly and slightly far from the deviation. Somehow, a false-positive can still be described as an anomaly. However, the percentage and accuracy are not much likeable and noticeable. For instance, the PPV value of Fig. 8(f) in the semi-structured experiment is 0.89. It has the lowest difference, with an overall mean value of 0.8983. Hence, the percentage of the object positively identified in Fig. 8(f) is almost 90%. Still, the image captured in Fig. 8(e) shows that the object has surpassed the frame edge at almost 40%. Based on the standard deviation calculation,  $\sigma$  the value is approximately 0.1, which is quite good for anomaly detection. It is because the range or frequency for an anomaly to happen is less and close to 0.

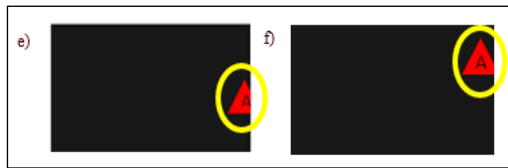


Fig. 8. True-positive anomaly versus False-Positive anomaly.

**B. Object Size**

The main objective of this experiment is to observe the pattern of mean and standard deviation values of each image taken from the vision sensor humanoid robot. With the pattern, it can be deduced on what condition the frames can be labelled as abnormal or normal. First, the colour and shape of the object are fixed throughout the experiment to ensure the output result is clear and easy to analyse. The object chosen is a cuboid and red. To control the experiment and for comparison, the mean and standard deviation value for the normal frame is recorded first. The recorded values are tabulated in Table II below. A normal scene is when no object is placed in the corresponding frame area.

TABLE II. MEAN AND STANDARD DEVIATION VALUES IN A NORMAL BEHAVIOR

Frame Number	Data	Value
Frame 1 (Center)	Mean ( $\times 10^2$ )	$2.02 \times 10^2$
	Std. Deviation ( $\times 10^1$ )	$4.06 \times 10^1$
Frame 2 (Left)	Mean ( $\times 10^2$ )	$2.20 \times 10^2$
	Std. Deviation ( $\times 10^1$ )	$3.68 \times 10^1$
Frame 3 (Right)	Mean ( $\times 10^2$ )	$2.06 \times 10^2$
	Std. Deviation ( $\times 10^1$ )	$3.94 \times 10^1$

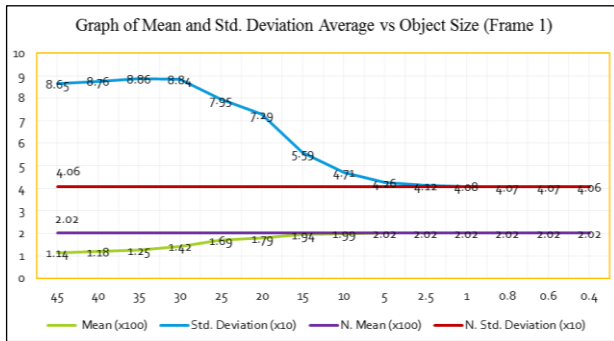


Fig. 9. Graph for mean and std. deviation average versus object size (Frame 1).

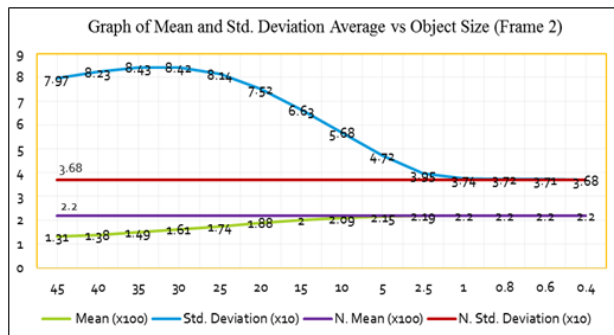


Fig. 10. Graph for mean and std. deviation average versus object size (frame 2).

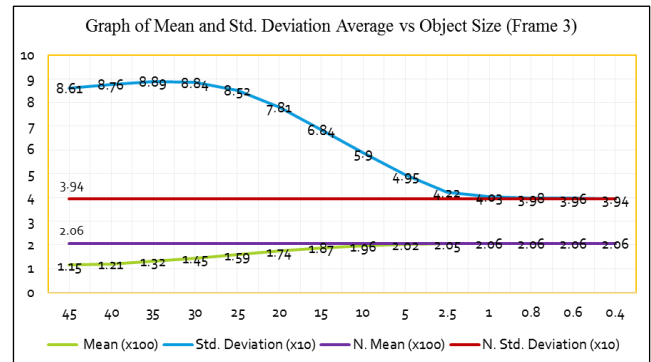


Fig. 11. Graph for mean and std. deviation average versus object size (Frame 3).

In Experiment B, only the object's size is manipulated. Other than that, the colour, shape and position are controlled during the experiment. The output data are plotted in the graph version, as shown in Fig. 9 to Fig. 11. The graphs show that the anomaly is identified when the cuboid size gets bigger than the typical size. The differences between the normal mean value and normal standard deviation value are very apparent too, with the biggest difference is recorded in Frame 3. The size for the cuboid reaches up to 50 cm x 50 cm and covers the whole frame wall. Moreover, anomalies also can be identified on the smaller object as long as the standard deviation value passes the threshold value. From the plotted graph above, cuboid size 1 cm x 1 cm is the minimum size detected as an anomaly in Frame 1. For Frame 2 and Frame 3, the smallest size is 0.8 cm x 0.8 cm. It is because when the object is smaller, the pixel value for the coloured object has no differences from the normal condition. Technically, it is considered normal behaviour, although the object is inside the frame. Therefore, it is known as a false-positive anomaly. More simply, pixel levels in the output image are too small and do not pass the normal frame value to be labelled as abnormal.

As for the bigger size, the pixel values for the coloured object are different from the plain frame as pixel intensity for the red colour is much higher than the grey one. This is also known as a true-positive anomaly.

**C. Location of Object**

In this experiment, the shape and colour of the cuboid are set not to change. However, the size is still different from the smallest to the biggest, which makes it possible to fit the cuboid size and calculate its position. The main objective of this experiment is to observe the location and position of the red cuboid image in a particular frame. With the pattern, it can be deduced which size of cuboid and position is suitable for humanoid robots to detect the anomalies. As stated earlier, the colour and shape of the object are fixed throughout the experiment to ensure the output result is clear and easy to analyze. For example, Frame 1 is still at the centre at 0° while Frame 2 is +40° and Frame 3 is -40°.

Here, Fig. 12 below shows the accomplished segmentation process to locate only the inside region of the object (dark grey colour) and eliminate the rest of the outer region (black colour).

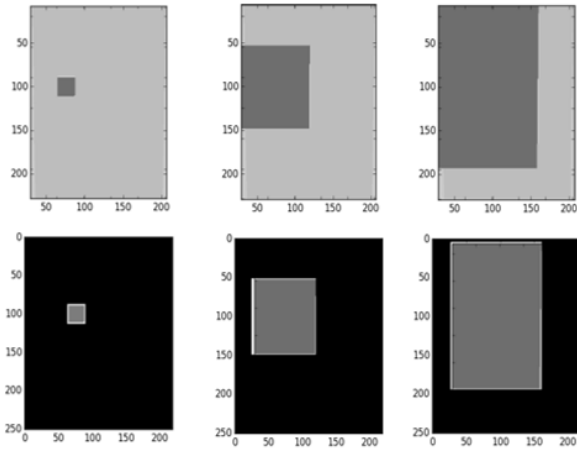


Fig. 12. Images captured by vision sensor for 5 cm x 5 cm, 20 cm x 20 cm and 35 cm x 35cm surface area.

Based on the calculation from each frame, it proves that the object's location from the humanoid robot is about  $8^{\circ}31'$  on the left side of the frame wall. The value is calculated based on Pythagoras' Theorem and Sine Rule.

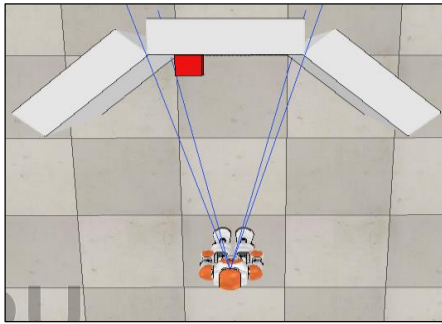


Fig. 13. Setting from the upper view between the wall frame and the robot to measure the angle.

The wall frame is 0.5 m in length. Starting from the centre until the left edge of the frame is 0.25 m. Then, the cuboid is set to be 0.1 m from the left side of the frame edge. So, the value of  $a = 0.15$  m. Value  $b$ , the distance between the robot and the frame, is fixed at 1 m. To calculate  $c$  and  $x$ :

$$c = \sqrt{b^2 + a^2} \quad (7)$$

$$c = \sqrt{1^2 + 0.15^2}$$

$$c = 1.01\text{m} \quad \text{and,}$$

$$\frac{\sin X}{0.15\text{m}} = \frac{\sin 90}{1.01\text{m}} \quad (8)$$

$$1.01\text{m} \times \sin X = 0.15\text{m} \times \sin 90$$

$$X = 8^{\circ}31'$$

So, the cuboid is located  $8^{\circ}31'$  between the centre and left of the robot. The outline can be seen in Fig. 13. Different situation with experiment C, the robot needs to locate the actual size and location of the anomalies on the frame wall after the anomaly behavior can be distinguished. Similar to Experiment B, the cuboid's colour, shape and position are fixed during the experiment. However, in this experiment, the

robot could spot the cuboid size location up to 20 cm x 20 cm for Frame 1 and Frame 3 and 25 cm x 25 cm for Frame 2. When the object is bigger than the limit size, the centre and edges of the object are not fully fitted in a particular frame. Therefore, the centre and edges of the cuboid are located differently and over the limit; even the standard deviation value differs from the normal behavior. This condition can be considered a true-negative anomaly. To make the robot detect bigger object than the limits, the size of the wall need to be bigger and extended. Using Pythagoras Theorem and Sine Rule, the actual location, right side and distance of the abnormal can be calculated. With this ability, anomaly detection can be improved as the robot can locate the object's location as long as it is still within the frame boundaries.

#### D. Position of Object

The main objective of this experiment is to observe the pattern of mean and standard deviation values in a particular frame despite the position not being fixed while the size is constant. With the pattern, it can be assumed that the best position is suitable for a humanoid robot to detect anomalies. From the results, the surface area of a 10 cm x 10 cm cuboid is chosen as the best measurement to get the median value. It also fits the normal frame and the other two slanted frames, regardless of various positions. As stated earlier, for making a comparison, the mean and standard deviation value for the normal frame is recorded first.

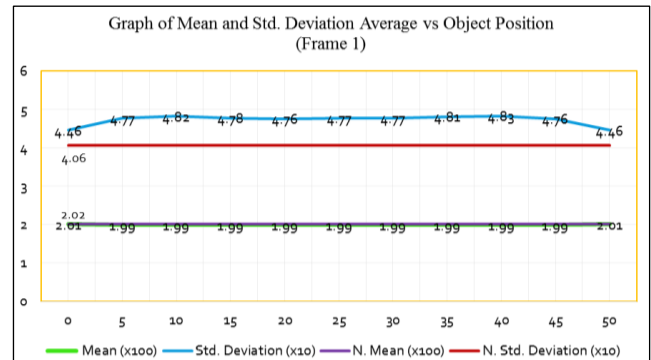


Fig. 14. Graph of mean and std. deviation average vs object position (frame 1).

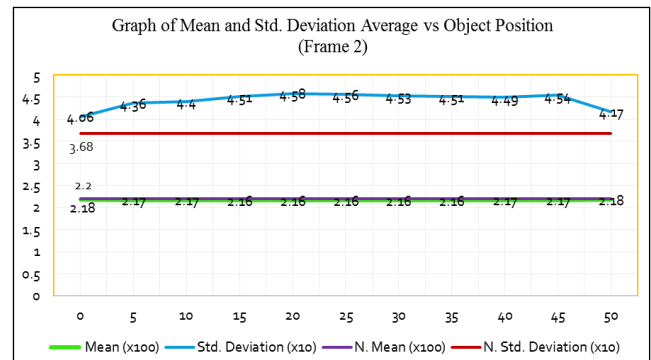


Fig. 15. Graph of mean and std. deviation average vs object position (frame 2).

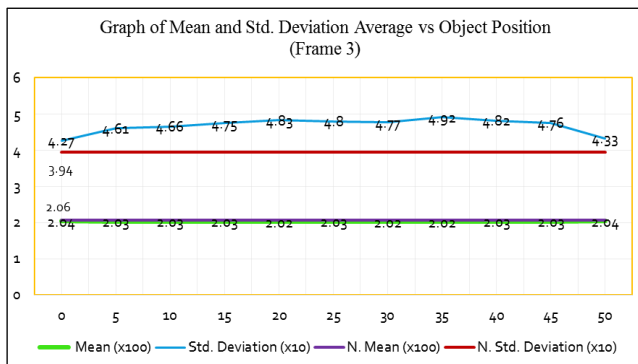


Fig. 16. Graph of mean and std. deviation average vs object position (frame 3).

In general, experiment D is quite similar to Experiment B. The changes in the scenario where the cuboid is placed at the halfway point but varies along different lengths on the frame. This experiment uses the output result in mean and standard deviation values to distinguish the anomaly pattern. A 10 cm x 10 cm cuboid was fixed during the experiment. The reason for using the mentioned size area is that it can be fitted on the surfaces of all kinds of frames. As mentioned in the prior subchapter, Frame 2 and Frame 3 were tilted ( $\pm 40^\circ$ ) from their original axis ( $\pm 0^\circ$ ). It also gives the best output image after being detected by robot vision. It has a median value with the highest and lowest mean and standard deviation pixels recorded from the previous experiment. The result shows that an anomaly can still be identified and evaluated, even though the standard deviation value has fluctuated in all frames. Because of the fixed size, the mean graph shown in Fig. 14 to Fig. 16 for each frame does not change critically. Their differences were at least 0.01 until 0.04 from the normal mean value.

## V. DISCUSSION

With all the observations, the calculated mean,  $\mu$  and standard deviation,  $\sigma$  for the initial and final point of the object in all three frames were recorded as the lowest. The biggest difference in  $\sigma$  value showed when the object moved nearer to the middle. It is because, at 0 cm and 50 cm, the cuboid is captured half of its actual size in frame boundaries. Thus the pixel intensity of the background value is the highest compared to the object colour value, making the standard deviation value drops significantly. The image pixel also shows the highest grey intensity when the object is located by the edge of any frame. All performance data in Experiments B, C and D are categorized as quantitative results.

Based on the graph trends corresponding to the tabulated data recorded in Experiment B, the  $\sigma$  curve ascends and achieves the highest peak when the object is 35 cm x 35 cm. In contrast, the mean line decreases slightly until its minimum value. It is because the pixel intensity for the red value surpasses the grey value, which gives great differences in deviation reading compared to the normal deviation data. However, it started to decline slightly after the highest peak as the pixel value for grey is untraceable inside the frame. It also happens when the object size is smaller. For example, when the measurement size is 1 cm x 1 cm, the line for standard deviation approaches the line for the normal value. The

changes happened in the mean line too. This can be explained as the pixel value for red being untraceable and considered a normal condition. This situation also applies in Frame 2 and Frame 3. Observing through each graph in Experiment D, the pattern for mean in anomaly and common condition shows little differences and is fixed horizontally. However, a different outlook is observed in the standard deviation pattern. The graph is up and down compared to the normal standard deviation line. The outcome shows that the standard deviation value is the best parameter to be observed when examining anomaly behavior and comparing them with the normal state.

Meanwhile, the camera on the humanoid robot followed the object no matter the target object's position in image segmentation. From the concept, the distance between the object on the frame and the NAO robot is always the same. Both images in the left and right eye are matched so that only the image of the target object is recognized, pinpointed and shaded in the bounding area. Interest points are grouped according to respective image position and disparity. Each of them carries a different pixel number.

However, the mean pixel number of a particular object is much alike, although it is located in various positions. To access each image's pixel value and calculate the mean and standard deviation, Open CV and Numpy plug-in library functions need to be explored. In this project, every image is set to produce only 3-bit pixel values. The pixels are red, green and blue intensity. The pixel value detected differed from what it was supposed to be whenever the object's position was between the edges of two frames, or the object surpassed the frame edge.

As a consequence, the segmented image produced different values of pixels. As stated earlier, anomaly behavior happens when there is unlike frequency happening at a time. Thus, when the mean, standard deviation, or PPV value produced differs from the normal state, it can be categorized as an anomaly, as stated by a study in [17, 18].

The proposition that can be made from Experiments A, B and D is that both  $\mu$  and  $\sigma$  values are reliable. The difference between normal and anomaly is comparable and can be proved based on the physical data and recorded image. To validate, each experiment is repeated five times to get accurate results, and the average value is calculated and recorded in tables. However, minor errors happen when the head yaw of the humanoid robot is rotated  $30^\circ$  to the left and right during detecting and scanning the frames repeatedly. Sometimes, the robot turned less than  $30^\circ$  or more than  $30^\circ$  from what it was supposed to. The effect of this the captured image is not similar as expected because the frame and point of vision are visibly not parallel because of the few internal factors such as joint errors and kinematic errors [19]. Because of that, the simulation must stop and take a break for 5 to 6 seconds. Then, it resumed again to get a precise result and eliminate the insignificant error that might change the results, as proposed by [20].

## VI. CONCLUSION AND RECOMMENDATIONS

In a nutshell, the main objectives of this project are successfully achieved using the semi-supervised approach to

learn the anomalies. PPV value, overall mean value, variance and standard deviation have been calculated to prove the anomaly statistically in that scenario. It also can be concluded that an object's positions greatly vary the standard deviation values. The location of the abnormal condition also has been proved during the experiment. It is proved by observing the maximum size of the object that can fit in the captured frame by ensuring all surfaces are within the range. Data obtained from Frame 2 differs from the data in Frame 3, although both slanted 40° from the horizontal. The setting for the frames and light tuning needs to be customized in the future. Besides that, a shadow effect is formed from different angles even though the light intensity for the surrounding has been set to default in the early setup. Furthermore, the calculation for mean and standard deviation pixel value can be made to access every bit image rather than only 3-bit colour to improve the study result. And lastly, the improvement could be focused on when there are ambient environmental changes in terms of illumination intensity, object colour and quantity of objects to be detected. The control system of the robot could also be improved by using closed loop estimation system [21] or by adaptive technique called disturbance observer [22].

#### ACKNOWLEDGMENT

The authors would like to acknowledge the funding supplied by Universiti Teknikal Malaysia Melaka (UTeM) under the Zamalah scheme to perform the study. The authors would also like to convey our credits and gratitude to the Robotics and Industrial Automation research group within the Faculty of Electrical Engineering.

#### REFERENCES

- [1] Zabalza, J., Fei, Z., Wong, C., Yan, Y., Mineo, C., Yang, E., Rodden, T., Mehnen, J., Pham, Q. C. and Ren, J., "Smart Sensing and Adaptive Reasoning for Enabling Industrial Robots with Interactive Human-Robot Capabilities in Dynamic Environments - A Case Study," *Sensors*, vol. 19, 2019. doi: 1354. 10.3390/s19061354.
- [2] Pasquale, G., Mar, T., Ciliberto, C., Rosasco, L., and Natale, L., "Enabling Depth-Driven Visual Attention on the iCub Humanoid Robot: Instructions for Use and New Perspectives," *Frontiers in Robotics and AI*, vol. 3, 2015. doi: 10.3389/frobt.2016.00035.
- [3] Leitner, J., Förster, A., and Schmidhuber, J., "Improving Robot Vision Models for Object Detection Through Interaction," *Proceedings of the International Joint Conference on Neural Networks*, 2014. doi: 10.1109/IJCNN.2014.5686837.
- [4] Flores, R., Karen, L., Trujillo-Romero, F. and Suleiman, W., "Object recognition modular system implementation in a service robotics context," pp. 1-6, 2017. doi: 10.1109/CONIELECOMP.2017.7891833.
- [5] Laielli, M., Smith, J., Biamby, G., Darrell, T., and Hartmann, B., "LabelAR: A Spatial Guidance Interface for Fast Computer Vision Image Collection," pp. 987-998, 2019. doi: 10.1145/3332165.3347927.
- [6] Pizzuto, G. and Cangelosi, A., "Exploring Deep Models for Comprehension of Deictic Gesture-Word Combinations in Cognitive Robotics," pp. 1-7, 2019. doi: 10.1109/IJCNN.2019.8852425.
- [7] Khairy, D., Salem A, Areed, M. F., Mohamed A. A. and Rania A. A., "An Algorithm for Providing Adaptive Behavior to Humanoid Robot in Oral Assessment", *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 9, 2022. doi: 10.14569/IJACSA.2022.01309119.
- [8] Martínez, E., del Pobil, Angel P., "Vision for Robust Robot Manipulation," *Sensors (Basel, Switzerland)*, vol. 19, 2019. doi:10.3390/s19071648.
- [9] Sanchez, T., Caramiaux, B., Thiel, P., and Mackay, W., "Deep Learning Uncertainty in Machine Teaching," pp. 173-190, 2019. doi: 10.1145/3490099.3511117.
- [10] L. Ma, M. Ghafarianzadeh, D. Coleman, N. Correll and G. Sibley, Simultaneous localization, mapping, and manipulation for unsupervised object discovery. In 2015 IEEE International Conference on Robotics and Automation (ICRA), (2015) 1344-1351. doi: 10.1109/ICRA.2015.7139365.
- [11] Zhou, Z. and Yatani, K., "Gesture-aware Interactive Machine Teaching with In-situ Object Annotations," 2022, pp. 1-14. doi: 10.1145/3526113.3545648.
- [12] Ari, D. and Alagoz, B. "A Review of Genetic Programming Popular Techniques, Fundamental Aspects, Software Tools and Applications. *Sakarya University Journal of Science*, 2021. doi: 10.16984/saufenbilder.793333.
- [13] B. Nuño, C. Carlos and H. Álvaro, "Improving the detection of robot anomalies by handling data irregularities," *Neurocomputing*, vol. 459, 2021, pp. 419-431. doi:10.1016/j.neucom.2020.05.101.
- [14] Gueye, T., Wang, Y., Mushtaq, R. T., Rehman, M., Ahmed, A., and Ali, H., "State of the art review on automatic sorting system for industrial robots using Internet of Robotic Things," 2022. doi: 10.21203/rs.3.rs-2329674/v1.
- [15] Espinosa, O., Castañeda, L., and Martínez, F., "Minimalist Artificial Eye for Autonomous Robots and Path Planning, 2016, pp. 232-238. doi: 10.1007/978-3-319-24834-9\_28.
- [16] Giovanni, M., Carmine, R., and Rezia, M., "Vision and Locomotion Control Systems in a bioinspired Humanoid Robot," *Proceedings of the Mediterranean Electrotechnical Conference - MELECON*, 2015. doi:10.1109/MELCON.2014.6820564.
- [17] Bucinkas, V., Dziedzickis, A., Sumanas, M., Sutynys, E., Petkevicius, S., Butkiene, J., Virzonis, D., and Morkvenaite-Vilkonciene, I., "Improving Industrial Robot Positioning Accuracy to the Microscale Using Machine Learning Method," *Machines*, vol. 10, no. 940, 2022. doi: 10.3390/machines10100940.
- [18] Chunkai, Z., Chen, Y., Yin, A., and Wang, X., "Anomaly detection in ECG based on trend symbolic aggregate approximation," *Mathematical Biosciences and Engineering*, vol. 16, pp. 2154-2167., 2019. doi:10.3934/mbe.2019105.
- [19] Keeskés, I., Odry, Á., and Odry, P., "Uncertainties in the Movement and Measurement of a Hexapod Robot". In: Awrejcewicz, J. (eds) *Perspectives in Dynamical Systems I: Mechatronics and Life Sciences*. Springer Proceedings in Mathematics & Statistics, vol. 362, 2019. doi: 10.1007/978-3-030-77306-9\_12.
- [20] Vocetka, M., Huňady, R., Hagara, M., Bobovský, Z., Kot, T., and Krys, V., "Influence of the Approach Direction on the Repeatability of an Industrial Robot," *Applied Sciences*, 2020. doi: 10. 8714. 10.3390/app10238714.
- [21] Mansor, N. N., Jamaluddin, M. H., and Shukor, A. Z., "Adaptive Control Technique Effects on Single Link Bilateral Articulated Robot Arm," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 7, pp. 512-520, 2021. doi: 10.14569/IJACSA.2021.0120759.
- [22] Aivaliotis, P., Papatitsa, E., Michalos, G., and Makris, S., "Identification of dynamic robot's parameters using physics-based simulation models for improving accuracy," *Procedia CIRP*, vol. 96, pp. 254-259, 2021. doi: 10.1016/j.procir.2021.01.083.

# Queueing Model based Dynamic Scalability for Containerized Cloud

Ankita Srivastava\*, Narander Kumar

Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India

**Abstract**—Cloud computing has become a growing technology and has received wide acceptance in the scientific community and large organizations like government and industry. Due to the highly complex nature of VM virtualization, lightweight containers have gained wide popularity, and techniques to provision the resources to these containers have drawn researchers towards themselves. The models or algorithms that provide dynamic scalability which meets the demand of high performance and QoS utilizing the minimum number of resources for the containerized cloud have been lacking in the literature. The dynamic scalability facilitates the cloud services in offering timely, on-demand, and computing resources having the characteristic of dynamic adjustment to the end users. The manuscript has presented a technique which has exploited the queuing model to perform the dynamic scalability and scale the virtual resources of the containers while reducing the finances and meeting up the user's Service Level Agreement (SLA). The paper aims in improving the usage of virtual resources and satisfy the SLA requirements in terms of response time, drop rate, system throughput, and the number of containers. The work has been simulated using Cloudsim and has been compared with the existing work and the analysis has shown that the proposed work has performed better.

**Keywords**—Cloud computing; scalability; containers; containerized cloud models; queueing model

## I. INTRODUCTION

Cloud computing has evolved into a highly dynamic computing model. It has gained attraction from various organizations due to its cost, availability, scalability, and security. It is an internet-based computing technology that provides higher-end computation and a shared pool of resources which are accessible on demand [1]. It has revolutionized the internet world through its hosting services and computational ability. Its unique technology has facilitated the user to pay for only those services and resources which have been demanded by them and further these resources can be increased and decreased depending upon the requirement. The potential and high capabilities have led to amplified productivity with reduced costs and flexibility as against the other IT industries [2]. The prime technology working behind the cloud is virtualization which enabled the cloud to instantiate various Virtual Machines (VMs) on one single physical machine (PM). Virtualization can occur at various levels like desktop, network, storage, and application [3]. It can affirm high performance, confidentiality, reliability, and security among VMs. One VM is isolated from the other VM on the same PM making it securely isolated. Despite various benefits exhibited through virtualization, applications demanding less isolation and maximum flexibility at runtime,

VM virtualization may not be sufficient enough to satisfy all the QoS standards [4]. The container-based virtualization is gaining more popularity these days because of the more dynamic and flexible nature of the workload which varies highly with time. It expedites the seamless movement of applications from one architecture to another as against the VMs virtualization. Container executes on a kernel with the equivalent performance as VMs but with lesser cost than expensive VM runtime management overhead [5]. Containers provide a good platform to execute microservices on the cloud and they provide good support for the technologies such as fog computing, and the Internet of Things (IoT) [6]. As container technology gained popularity various large-scale IT industries providing cloud services have come up with their container-based cloud services.

The most renowned service models available in the cloud are Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) with various energy efficient datacenters (DC) which are solely responsible for managing the scalability through resource management and load optimization [7]. With the PaaS service, the users can deploy any applications on the cloud. This model encapsulates the underlying infrastructure and facilitates the user to deploy the applications anywhere without giving a single thought about infrastructure management. One of the components of the PaaS service is containers and they are its enablers [8]. So, a user application can be deployed on a single cloud infrastructure as a unique block or deployed separately in different cloud infrastructures.

The key characteristic of the cloud to scale up has attracted a lot of users. The variation and fluctuation in workload have compelled the cloud providers to scale up the resources (VMs or containers) dynamically as per the requirement. The cloud has eased the process of obtaining and releasing resources but it can be challenging to decide how many resources are needed to handle a fluctuating workload. There is an urgent demand for a model which can provision and de-provision the resources dynamically at the burst of demands. Despite the development of container technology and harnessing its potential, there is still room for the improvement in dynamic scaling of cloud resources. Insufficient scalability which is not competent enough to confront the variation in the workload intensity may lead to under-provisioning (UP) or over-provisioning (OP) of the resources. In the UP scenario, the performance of the cloud degrades and SLA is violated. While in OP there is low consumption of resources that are allocated resulting in a higher cost for the providers. As a result, in response to dynamic changes in the global arrival rate during

\*Corresponding Author.

runtime, adaptation mechanisms are covered for polished dynamic scalability. Appropriate dynamic scalability is the demand of the time and it affirms the performance of the SLA while making the cost low. An efficient technique for dynamic scalability is required for fulfilling the requirement of both the users and CSP.

This work has proposed a dynamic scaling approach for the containerized cloud. The approach enables us to acquire the dynamic and scalable nature of cloud computing and analyze its efficaciousness. The model tries to estimate the future resource demand and provision the resources in a dynamic way for mitigating the SLA violation and reducing the cost incurred by the system. The work is simulated in Cloudsim and the work is evaluated under the various quantity of workload. The main contribution of the paper goes as:

- A queuing model is proposed to estimate and acquire the behavior of containerized DC.
- The load balancer model and container model are discussed.
- The mathematical formulation has been derived from the analytical model for various QoS measures.
- Simulation of the work is performed on Cloudsim.

The remaining paper is compiled as follows: the literature study associated with the work is done in Section II. The proposed work is discussed in Section III. Section IV performs the results and discussion and lastly, the work is concluded in Section V.

## II. RELATED WORK

Containerization is not a novice concept of computer science. It was existing back in 1972 on Linux or Unix systems in different ways [9]. It aided the developer in providing an efficacious programming environment which has a quite reduced operational cost. Docker has adopted container technology and led to the start of open containers in the industries like Google, Microsoft, and many more and it is getting popularity day by day due to its isolation strategy. Fig. 1 describes the container in the cloud system with its private OS, interface, and file system. Cloud containers provide a thin encapsulation over the application so its deployment is relatively easier and faster. Initially, it started with the VMs which are light. These technologies possessed an isolated OS on which the application can be deployed [10]. Containers have several benefits over VMs [11]. Firstly, compared to virtual machines, containers use host system resources far more efficiently. Second, starting and stopping the containers only takes a minute time. Next, the mobility container prevents inter-system dependency conflict and guarantees its separate functioning from the system on which it is hosted. Fourth, unlike VMs, which are frequently not distributed production environments, containers possess the feature of being exceedingly lightweight, enabling end users to operate dozens or more of them simultaneously. Fifth, instead of having to go through hours-long installation and configuration hassles, end users of apps can instantly download and run sophisticated software. Additionally, unlike virtual machines (VMs), which strive to virtualize an external

environment, a container's primary goal is to make an application fully portable and independent [12].

The containerized cloud is emerging as one of the most challenging issues over the past few years and a lot of work has been published in this regard. This section studies the relevant work associated with the scalability and performance of container-based cloud models. Some study is associated with scalability to provide better insight into scalability and some shows the scalability in container clouds.

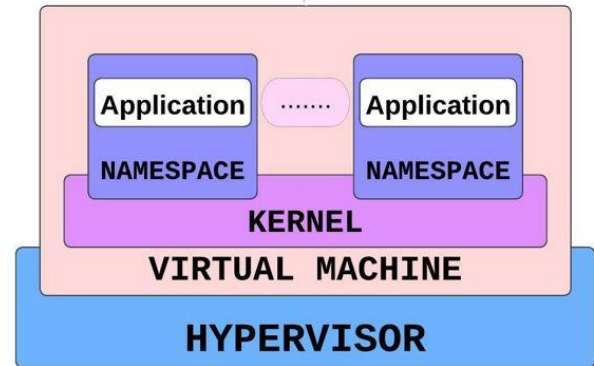


Fig. 1. Container structure.

Scalability has once been qualified as the key feature contributing to the efficient working of cloud-based services. A deep scaling methodology has been introduced in [13] where three components have been included for effective resource utilization. First, it forecasted the workload then it mapped the workload intensity to the approximated CPU utilization and lastly, an auto scale method is developed for maximizing the CPU utilization. A proactive elastic model was defined in [14] which resolved the scalability issues in cloud-based IoT systems. It utilizes the ant colony optimization technique along with the Markov chain for scheduling the resources efficiently which enhances the performance and maintains the QoS measures. It improved the response time and request throughput. A benchmarking method is proposed that defined a framework for scalability benchmarking tools for quantifying the scalability. It included the scalability metrics and measurement methods to specify the achievement of the given service level objectives. It also provided the facility for configuring the scalability parameters for getting an efficient response [15]. The author in [16] proposed a container-based autoscaling procedure that used a heuristic technique for utilizing the resources efficiently. It improved the execution time, throughput, response time, and the minimum number of containers. The author in [17] addresses the two major scalability metrics volume scalability and quality scalability. Volume scalability is highly influenced by the scaling of service volume while quality scalability is affected by the service quality provisioned. These parameters quantify the technical scalability and helped in assessing the impact of demand on the service. Besides, they also aided in designing and performing scalability testing with the motive of the identification of the components that affect the scalability performance.

Scalability in container clouds has made the processing of cloud applications lightweight and efficient. An automatic scaling method is discussed in [18] where it reduces the response time, energy consumption, and better CPU utilization. An analytical model based on the stochastic technique for the container-based DC has been discussed in [19]. It studied and analyzed the performance of the cloud system with respect to mean job delay and job rejection probability. It created a framework for container emulation and assessed the same against the suggested stochastic technique. Through experimental development, the suggested model is validated using actual data. Insight into DC planning is provided to system designers by numerical verification. Another approach is introduced in [20] in which AWS autoscaling is implemented which facilitate estimating the future workload. It applied a future prediction algorithm using Prophet API. It studied the CPU utilization and the creation of new EC2 instances when the workload is heavy. An auto-scaler-based model has been discussed in [21] which provide the architecture for the container-based application. It has included a monitoring mechanism, prediction model, time series model, and decision mechanism. The prediction utilized the time series to predict the future workload. It has provided better provisioning and speedy elasticity. The author in [22] introduced a framework for auto-scaled containerized applications which is governed by workload demand. It offered both reactive and proactive scaling. Reactive scaling was implemented using the threshold rules and proactive scaling utilized a neural network. It ensured the requirement of QoS. Another container-based module was developed in [23] which provided efficient provisioning. It used an adaptive function tree for scalable container provisioning. It mitigated the provisioning cost further by using a fetching mechanism showing the quality of on-demand and I/O efficiency. It turned out to be providing better scaling, response time, and provisioning. A horizontal scaling technique is discussed in [24] which configured the services in a docker container while the workload was balanced using the load balancer. It calibrated the infrastructures depending on the number of predicted users. It expanded the infrastructure and processing capability in a short duration and offered a fault-tolerant system for medium and small-scale industries. Another technique for resource utilization in a cloud-based application is discussed in [25] for container clouds leveraging the vertical elasticity of Docker. The resource coordinator and monitoring policies are implemented during the execution of tasks. Scalability parameters are the configurable parameters in the procedure.

To the greatest of our knowledge and as of this time, there hasn't been any research available for the effectiveness and dynamic scalability of containers published in the literature. The existing work does not consider the dynamic scalability in the containers which has provided a cost-effective solution to the virtualization. It is of utmost importance to identify the number of containers required to cope with the highly dynamic workload to satisfy the SLA and QoS requirements. Dynamic scalability is attained only when there is neither overprovisioning nor under-provisioning. Overprovisioning may result in higher costs as more containers will be

engrossed while under-provisioning leads to SLA violation. Therefore, the main distinction between our study and the studies listed above is that in addition to forecasting workload, we also forecast the future need for computing resources. Furthermore, in contrast to most techniques that focus on only one factor (CPU utilization), our model provides cloud providers with more information about the timely scaling and descaling of containers' and VMs' volume. This not only decreases the cost incurred by the users but also improves the user's experience and also mitigates the financial burden of service provider and infrastructure cost due to the efficient and wise usage of the resources.

### III. PROPOSED WORK

#### A. Problem Formulation

A model consisting of DC consists of PMs which has the capability of holding various VMs which are further profound enough to hold various containers representing the real practical scenario of current existing cloud services. A hypervisor is held responsible for allocating various VMs to a PM while multiple containers can be allocated to a VM. The task execution request raised from the different users is being sent to the load balancer (LB). This LB routes the traffic to the PMs for execution. These requests are sent to a buffer system which is linked to the LB queue from where it is sent to the containers for the allocation of the resources and their execution. The tasks from the queue are allocated to the containers as per the availability of the resources. Whenever the user demands a new container with a particular requirement of the resources, the establishment of SLA between the final users and the CSP is agreed upon by the delivery of the requested QoS. If the breach in the agreed SLA happens the CSP is supposed to pay the penalty to consumers. The flow of the tasks happens as end users put in the request and it is sent to the LB. This LB receives the requests and distributes the tasks to the PMs as per the allocation policy utilized. Each task is allocated a unique container. As the task request increases the VM scales up the through the addition of the container for the execution of the request. Mostly the companies utilizing the features of the famous company Docker [26], use at least 18 containers simultaneously. Let us consider the  $m$  PMs which are represented as  $P = \{P_1, P_2, P_3, \dots, P_m\}$  which can hold a maximum of up to  $n$  VMs represented as  $V = \{V_1, V_2, V_3, \dots, V_n\}$ . A VM can contain maximum  $l$  containers represented as  $C = \{C_1, C_2, C_3, \dots, C_l\}$ . So, a PM can be scaled maximum to  $n$  VMs which can accommodate a maximum of  $n \times c$  containers.

#### B. Queuing Model

The PMs undertaken in the DC have a similar configuration. The requests generated by the end users are sent to the queue and are served at each node in the DC based on a first come first serve basis. When the requests are executed and served well then they exit from the system. This paper has assumed that each request is served in only one container and one container will serve only one request. The DC is modeled using the open Jackson queuing model represented in Fig. 2.



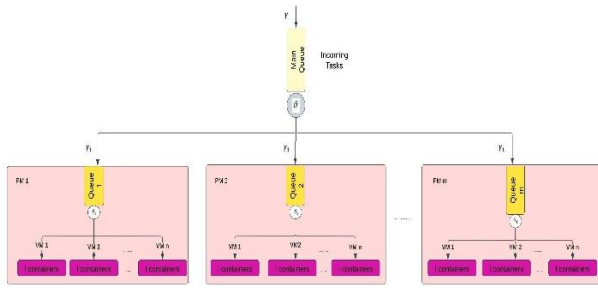


Fig. 2. Queuing model.

### C. Load Balancer Model

The LB is held responsible for managing the huge load which comes with cloud computing in the form of an ample number of requests from the end users. To serve the requests well the LB is modeled as an M/M/1 queuing model having the provision of the infinite capacity task requests buffer and the arrival of requests is supposed to be one by one [27]. The Markov chain with the continuous time of the LB model, and state  $t$  depicts the task number which means  $t - 1$  tasks are waiting to get allocated in the queue. The arrival of the tasks happens at the  $\gamma$  rate similar to that of the Poisson procedure in which the arrival duration of two immediate tasks is independent and the distribution is exponential according to the rate  $1/\gamma$ . The serving time to the task at the PM in the LB is exponentially distributed over the  $\vartheta$  rate and  $1/\vartheta$  is the mean serving time. If  $\alpha < 1$  the M/M/1 is assumed to be stationary, where  $\alpha = \frac{\gamma}{\vartheta}$ . Let the probability be  $\pi_t$  be for the  $t^{th}$  state. The following equations can be summated utilizing the balanced equation [28]:

$$\gamma\pi_0 = \vartheta\pi_1 \quad (1)$$

$$(\gamma + \vartheta)\pi_t = \gamma\pi_{t-1} + \vartheta\pi_{t+1}, \quad t \geq 1 \quad (2)$$

From equation 1 and 2, it can be written

$$\pi_t = \alpha^t \pi_0, \quad t \geq 1 \quad (3)$$

According to the normalized equation,

$$\sum_{t=0}^{\infty} \pi_t = 1 \quad (4)$$

One can deduce that,

$$\pi_0 = 1 - \alpha \quad (5)$$

And then the steady-state probability of  $t$  tasks in the queue can be given as:

$$\pi_t = (1 - \alpha)\alpha^t, \quad t \geq 0 \quad (6)$$

The number of tasks on an average queued in LB can be deduced as:

$$U_{lb} = \sum_{t=0}^{\infty} t\pi_t = \alpha(1 - \alpha) \sum_{t=0}^{\infty} t\alpha^{t-1} = \frac{\alpha}{1 - \alpha} = \frac{\gamma}{\vartheta - \gamma} \quad (7)$$

The average response time that the tasks in the queue obtained can be evaluated through Little's law [29] given as:

$$R_{lb} = \frac{U}{\gamma} = \frac{1}{\vartheta - \gamma} \quad (8)$$

### D. Container Model

The paper has considered a DC containing various PMs having various VMs designated to hold one or more container instances executing on it. The local Scheduler (LS) and the runtime component (VMs) are the two major holdings of a PM. Fig. 3 demonstrate the placement of VMs and containers and LS in a PM. The container is executed on these VMs as an isolated thread in a similar namespace with a guest OS shared among other containers in the same VM. The hypervisor performs the operations that include resource management for placing the containers in the pool of VMs in accordance with the workload being requested from the users.

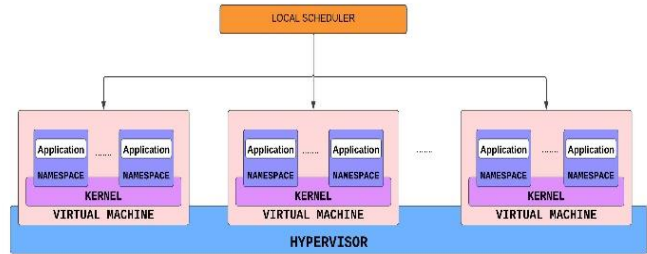


Fig. 3. PM structure.

Let's consider CDC contains  $m$  PMs with LS modeled as M/M/1/C model [30] with  $n$  VMs and  $l$  containers  $\eta = n \times l$  and thus making the queue full. This implies that PM is exhausted with the resources and is not in a condition to accept any new task until it gets finished up with the tasks previously allotted to it. So, it will reject the incoming new tasks. According to Burke [31], the departure procedure in the queue M/M/1 follows the Poisson process with the same rate  $\gamma$ . Thus, the tasks arriving at each PM follow the Poisson rate with  $\gamma_1 = \gamma/m$  and each task is served with a service time exponentially distributed with an average  $1/\vartheta_1$ . As the queue is finite in size so, for all the values of  $\gamma_1$  and  $\vartheta_1$  the system is stable. The  $i^{th}$  PM with  $t$  tasks in the queue has an equilibrium probability that can be defined as:

$$\pi_i^t = \begin{cases} \frac{(1-\omega)\omega^t}{(1-\omega^{\eta+1})}, & \omega \neq 1 \\ \frac{1}{\eta+1}, & \omega = 1 \end{cases} \quad (9)$$

The rate at which tasks are lost at the  $i^{th}$  PM at the LS queue can be obtained as:

$$L_i = \gamma_1 \pi_i^\eta = \gamma_1 \frac{(1-\omega)\omega^\eta}{(1-\omega^{\eta+1})} \quad (10)$$

The LS queue has  $i^{th}$  PM whose throughput is given as:

$$N_i = \gamma_1(1 - L_i) = \gamma_1 \frac{1-\omega^\eta}{1-\omega^{\eta+1}}, \quad \omega \neq 1 \quad (11)$$

Similarly, the volume of tasks available in  $i^{th}$  PM at the queue is:

$$D_i = \sum_t^\eta t\pi_i^t = \begin{cases} \frac{\omega}{1-\omega} \times \frac{1-(\eta+1)\omega^\eta + \eta\omega^{\eta+1}}{1-\omega^{\eta+1}}, & \omega \neq 1 \\ \frac{\eta}{2}, & \omega = 1 \end{cases} \quad (12)$$

The number of tasks undergoing the service is:

$$S_i = 1 - \frac{1-\omega}{1-\omega^{\eta+1}}, \omega \neq 1 \quad (13)$$

So, the tasks waiting in the queue can be given as:

$$W_i = D_i - S_i \quad (14)$$

The CPU utilization can be given as:

$$U = \frac{N_i}{\vartheta_1} \quad (15)$$

The waiting time for the tasks at  $i^{th}$  PM is:

$$WT_j = \frac{W_i}{\gamma_1} \quad (16)$$

Thus, the response time at  $i^{th}$  PM is evaluated as:

$$R_i = \frac{D_i}{N_i} = \frac{1}{\vartheta_1 - 1} - \frac{\eta\gamma_1^\eta}{\vartheta_1^{\eta+1} - \gamma_1^\eta}, \omega \neq 1 \quad (17)$$

For the later part of PM, each VM is modeled as the servers with  $l$  servers and the queue is not available with these servers i.e., M/M/1 [32] where  $l$  depicts the volume of containers available with each VM. As stated in [31], each virtual machine's incoming tasks follow a Poisson process with a rate of  $\gamma_2$ , indicating that each container receives an equal amount of requests. Since there are  $n$  VMs so each will get the tasks with an arrival rate  $\gamma_2 = \gamma_1/n$ . This will provide a balance system as each VM has a similar configuration. The service rate of each container can be taken as  $\vartheta_2$ . The task incoming at the VM can be visualized as a birth and death process. In a state  $a < l$ , the rate of the incoming task is  $\gamma_a = \gamma_2$  where  $\gamma_l = 0$ . While in the state  $a = 0, 1, 2 \dots l$ , the death rate  $\vartheta_a = a\vartheta_2$ . Let  $s_a$  be the stationary probability with  $a$  tasks in the  $a^{th}$  VM. It is observed:  $\gamma_2 s_{a-1} = a\mu_2 s_a$ , ( $0 \leq a \leq l$ ), from the local balance equation. With  $\delta = \gamma_2/\vartheta_2$ , it can be written as:

$$s_a = s_{a-1} \frac{\delta}{a} = s_0 \frac{\delta^a}{a!}, 0 \leq a \leq l \quad (18)$$

After the application of standardization condition [33],  $s_0$  can be generalized as:

$$s_0 = \frac{1}{\sum_{a=0}^l \frac{\delta^a}{a!}} \quad (19)$$

It can be further deduced:

$$s_a = \frac{\delta^a}{a!} \frac{1}{\sum_{a=0}^l \frac{\delta^a}{a!}}, 0 \leq a \leq l \quad (20)$$

The loss probability  $L_a$  for the tasks lost at  $a^{th}$  VM, as the VM was full, is recognized as:

$$L_a = \frac{\frac{\delta^l}{l!}}{\sum_{a=0}^l \frac{\delta^a}{a!}} \quad (21)$$

As there is no queue for the VM, so the tasks' volume in the VM

$$D_a = \delta(1 - s_a^{loss}) \quad (22)$$

As earlier, the response time is evaluated as:

$$R_a = \frac{1}{\vartheta_2} \quad (23)$$

As it is already known, when the LB sends the request, a job can only be carried out by one PM and in one VM by a container. Thus, the response time of the tasks before they went for execution can be summed as:

$$R = R_{lb} + R_i + R_a \quad (24)$$

With a similar analysis, the task being rejected in DC is:

$$L = L_a + L_i \quad (25)$$

#### IV. RESULTS AND DISCUSSION

The model proposed above is simulated through a series of experiments to analyze its effectiveness. The simulation has been performed on a personal computer with a 2.30 GHz Intel Core i3 processor and 4GB of RAM. The simulation tool used is Cloudsim. Initially, the DC is configured with 5 PMs and each PM is capable of supporting 10 VMs which varies to 50 VMs while each VM can accommodate a maximum of up to 18 containers. The arrival rate of the task varies from 1000 to 10,000 tasks per second. The task in the queue requests for execution which is serviced in 0.0001 seconds. The maximum capacity of the queue is 300. The LS service the request on an average of 0.001 seconds. The experiment is performed with 100 repetitions for efficient analysis.

##### A. Response Time

The response time of the system is very much affected by the volume of VMs which is analyzed with the varied task arrival rate. Fig. 4 illustrates the same. Here, the capacity of each VM to hold containers is 20. From the figure, it can be observed that the increment in response time to the task arrival is quite proportional. It is analyzed, for all the given scales of VMs, there are no substantial change in the response time when the arrival rate of the task varies from 5500 tasks per second to 9500 tasks per second. As the tasks arrival rate increases from 9500 tasks/second the response time increase exponentially for all the scales of containers. The response time 0.41second is observed in the 20 VMs scenario when the arrival rate is 10,000. The proper configuration of the VM can be chosen if the minimum response time is one of the QoS targets to be achieved for SLA.

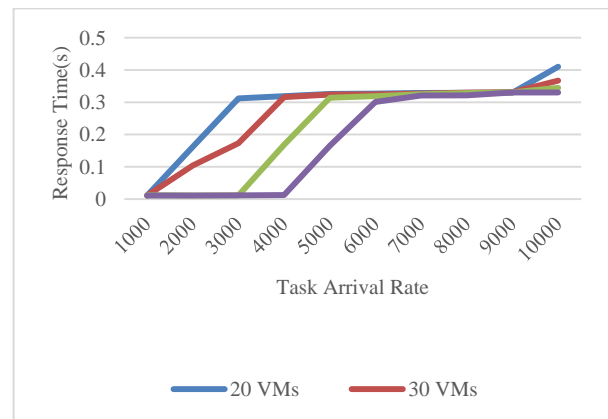


Fig. 4. Response time.

**B. Drop Rate**

The drop rate of the system is defined as the rate at which the tasks are dropped or rejected because of either lack of space in the LS queue or a lack of capacity in DC. Fig. 5 depicts the drop rate against the task arrival rate with the varied number of containers. Each VM has 18 containers. It can be observed from the figure that initially there is not much drop in the tasks but as the task rate increases the drop rate increases. This increase varies differently with a different configuration. In the case of 20 VMs, the loss starts after the 2000 task arrival rate is reached while in the 50 VMs case this loss starts after 5000 tasks/sec. Until the rate reaches 5000 the 50VMs configuration doesn't show any loss while the 3015 tasks/s are lost in the same task arrival rate as the 20VMs configuration. It can be deduced with the increased number of containers there is less loss of tasks.

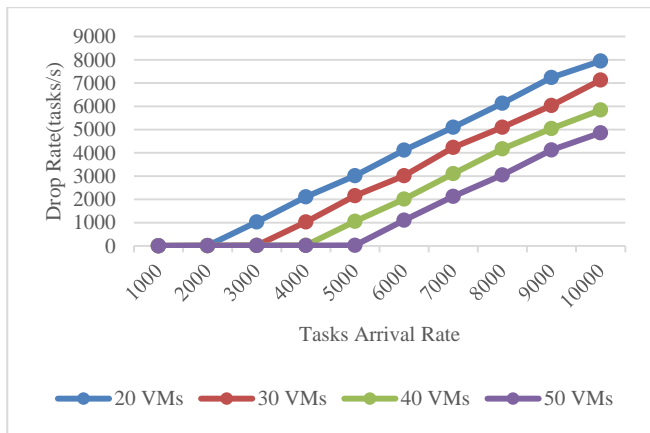


Fig. 5. Drop rate.

**C. Throughput**

The system throughput is being analyzed against the task arrival rate with all four configurations of the containers. Fig. 6 shows the variation of the system throughput measured in tasks per second. As it can be observed that in all four cases the system throughput is similar till 2000 tasks/sec. The impact of the different configurations of containers can be

seen beyond 2000 tasks/sec. The system performs better with a large number of containers. There is not much variation that can be seen when the rate of the task reaches 2000 tasks/sec in the first case, 3000 tasks/sec in the second case, 4000 tasks/sec in the third case, and 5000 tasks/sec in the last case. After a certain threshold, the throughput has become quite fixed. The requirement of predefined system throughput in SLA can be resolved using the selection of the best configuration of containers by the service providers.

**D. Number of Containers**

To study the effects of the number of containers on response time and drop rate the tasks arrival rate has been fixed at 9000 tasks/sec. The number of containers has been increased from 8 to 18 containers. Table I represents the response time. It is observed from the table as the number of containers increases the response time decreases. A dramatic decrease can be seen after the 16<sup>th</sup> container. The response time of the system is highly dependent on the volume of containers. The system drop rate is also getting highly influenced by the number of containers. It can be analyzed that as the containers increase the drop rate decreases. It shows that to keep the drop rate below 3010 tasks/sec the minimum number of VMs and containers is 50 and 8 respectively.

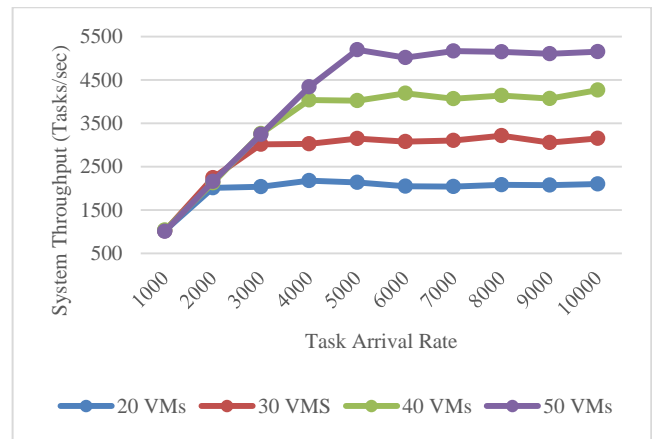


Fig. 6. System throughput.

TABLE I. SYSTEM RESPONSE TIME AND DROP RATE WITH RESPECT TO THE NUMBER OF CONTAINERS

No. of Containers	System Response Time				System Drop Rate			
	20	30	40	50	20	30	40	50
8	0.317	0.311	0.293	0.284	6135	5147	4087	3010
9	0.316	0.308	0.289	0.282	5948	5084	3942	2985
10	0.3149	0.304	0.286	0.2815	5827	4972	3875	2875
11	0.315	0.3037	0.288	0.281	5773	4864	3751	2870
12	0.3157	0.3021	0.287	0.2807	5648	4784	3617	2861
13	0.3154	0.2998	0.285	0.2794	5584	4743	3561	2756
14	0.3148	0.2994	0.283	0.279	5538	4476	3548	2641
15	0.297	0.287	0.277	0.274	5416	4507	3472	2571
16	0.283	0.278	0.264	0.258	5386	4459	3378	2468
17	0.257	0.246	0.238	0.2334	5258	4319	3307	2402
18	0.226	0.218	0.210	0.204	5156	4238	3193	2354

### E. Comparison with other Algorithms

The response time and system drop rate are compared with the existing work [11] and [16] for the 50VMs with 18 containers each. From Fig. 7, it can be observed that till the 4000 tasks/sec there is not much variation among the algorithms. As the rate increases the proposed shows better results. With 10000 tasks/sec, the response rate is 0.3301sec and that of [11a] is 0.601 sec. From Fig. 8., it can be deduced that till the task rate is 4000 all the algorithms show the same drop rate but as the task rate increases there is an exponential increment in the drop rate. With 10000 tasks/sec, the drop rate of the proposed algorithm is 4859 tasks/sec and while that of others is 5338 and 5812 respectively. The suggested method is significantly more effective than others.

The results obtained above have demonstrated that the increment in task arrival rate affects the QoS measures depending on the containers available in the DC. So, it's very essential to scale up or scale down the container instances depending upon the rate of the incoming task. In addition to this, the number of containers available has to fulfill the SLA requirements. Besides, in the DC the workload is very dynamic and to provision, the minimum containers dynamically which can fully satisfy the SLA requisite which monitors the usage of virtual resources and modify the number of resources to be used is of utmost importance. Therefore, the main challenge that needs to be worked upon is the engagement of the minimal number of containers for fulfilling the SLA exigencies. Allocation of a greater number of containers than required may lead to the OP which increases the cost. Deploying a lesser volume of containers than expected may result in UP leading to more SLA violations. Therefore, dynamic scalability is the requirement of the time to avoid the situation of under and over-provisioning. The proposed algorithm facilitates the service provider to identify the minimal containers required while the rate of the task fluctuates helping in scaling up and down the resources and maintaining the SLA.

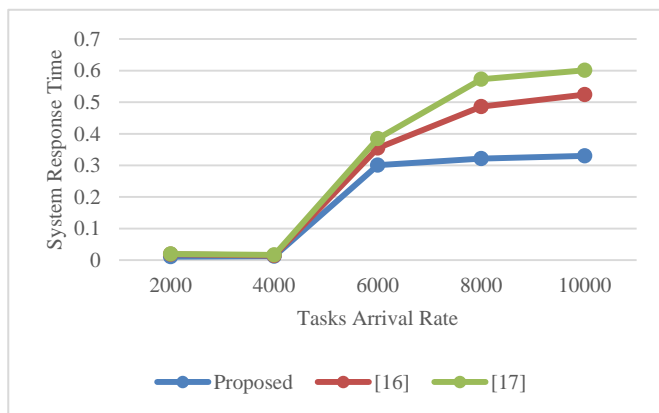


Fig. 7. Comparative analysis of system response rate.

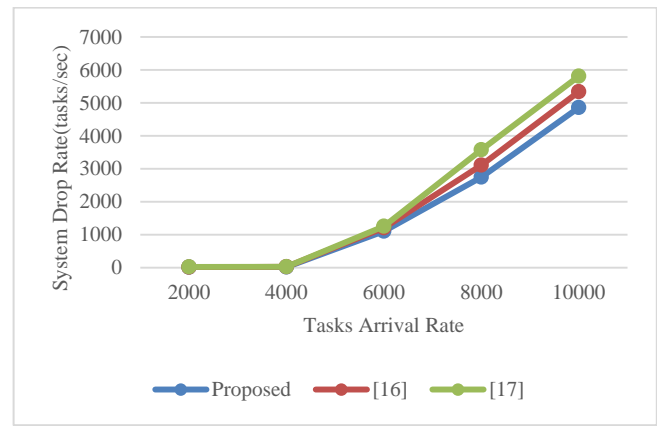


Fig. 8. Comparative analysis of system drop rate.

Further, the resources are allocated to the tasks in the order of their arrival. It may not consider the priority tasks which can be handled in further study. Since the arrival rate of the tasks is considered fixed which may differ in real life scenario as the arrival rate can vary with the state and thus making the potential customer to switch other service due to long waiting queue and thus it may affect the efficiency of the system.

### V. CONCLUSION

This paper has proposed a queuing model for dynamic scalability in containerized clouds to analyze the workload and the effects of scaling on the QoS parameters. It also suggests the number of containers is scaled up or down for the requirement of a particular given SLA. A mathematical model is developed for identifying the key performance metrics. The model predicts and approximates the resource request for future requirements to mitigate the SLA violations and provide cost-effective solutions. The proposed methodology can also be used to scale the DC containers to guarantee the QoS parameters. The model is proficient enough in deciding the number of containers required for the provision or deprovisioned as per the given workload situation to meet up the SLA demand and QoS metrics. The proposed model is tested against some existing work and has turned out to be performing better. In future work, the model can be implemented in a real working environment and more SLA parameters can be included for the analysis. Further, clustering technique can be included and the model can have queue classified according to the requirements of the tasks like some tasks may require more processing units while some require more storage unit.

### REFERENCES

- [1] Amini Motlagh, A., Movaghar, A., & Rahmani, A. M., "Task scheduling mechanisms in cloud computing: A systematic review," International Journal of Communication Systems, vol. 33, no. 6, pp. e4302, 2020. <https://doi.org/10.1002/dac.4302>.
- [2] V. Eramo, F. G. Lavacca, T. Catena, and P.J. Perez Salazar, "Proposal and investigation of an artificial intelligence (AI)-based cloud resource allocation algorithm in network function virtualization architectures," Future Internet, vol. 12, no. 11, pp. 196, 2020. <https://doi.org/10.3390/fi12110196>.

- [3] A. Bhardwaj, and C. R. Krishna, "Virtualization in cloud computing: Moving from hypervisor to containerization—a survey," *Arabian Journal for Science and Engineering*, vol. 46, no. 9, pp. 8585-8601, 2021. <https://doi.org/10.1007/s13369-021-05553-3>.
- [4] B. Varghese, R. Buyya, "Next generation cloud computing: new trends and research directions," *Future Gener. Computer Syst.* vol. 79, pp. 849–861, 2018. <https://doi.org/10.1016/j.future.2017.09.020>.
- [5] H. Khazaei, C. Barna, N. Beigi-Mohammadi, M. Litoiu, "Efficiency analysis of provisioning microservices," *International Conference on Cloud Computing Technology and Science (CloudCom)*, IEEE, pp. 261–268, 2016. <https://doi.org/10.1109/CloudCom.2016.0051>.
- [6] P. Di Francesco, P. Lago, and I. Malavolta, "Architecting with microservices: A systematic mapping study," *Journal of Systems and Software*, vol. 150, pp. 77-97, 2019.
- [7] Y. Saadi, S. El Kafhali, "Energy-efcient strategy for virtual machine consolidation in cloud environment," *Soft. Comput.* vol.24, no.19, pp. 14845-14859, 2020. <https://doi.org/10.1016/j.jss.2019.01.001>.
- [8] I. Kabashkin, "Availability of applications in container-based cloud PaaS architecture," In *International Conference on Reliability and Statistics in Transportation and Communication*, Springer, pp. 241-248, 2018. [https://doi.org/10.1007/978-3-030-12450-2\\_22](https://doi.org/10.1007/978-3-030-12450-2_22).
- [9] J.P. Martin, A. Kandasamy, K. Chandrasekaran, "Exploring the support for high performance applications in the container runtime environment," *Human-Centric Comput. Inf. Sci.*, vol. 8, no.1, pp. 1–15, 2018. <https://doi.org/10.1186/s13673-017-0124-3>.
- [10] L.Cai, Y. Qi, W. Wei, J. Li, "Improving resource usages of containers through auto-tuning container resource parameters," *IEEE Access*, vol. 7, pp. 108530–108541, 2019. <https://doi.org/10.1109/ACCESS.2019.2927279>.
- [11] A. Bhardwaj, and C. R. Krishna, "Virtualization in cloud computing: Moving from hypervisor to containerization—a survey" *Arabian Journal for Science and Engineering*, vol. 46, no. 9, pp. 8585-8601, 2021. <https://doi.org/10.1007/s13369-021-05553-3>.
- [12] B. Tan, H. Ma, Y. Mei, and M. Zhang, "A cooperative coevolution genetic programming hyper-heuristic approach for on-line resource allocation in container-based clouds," *IEEE Transactions on Cloud Computing*, vol.10, no.3, pp. 1500-1514, 2020. <https://doi.org/10.1109/TCC.2020.3026338>.
- [13] Z. Wang, S. Zhu, J. Li, W. Jiang, K. K. Ramakrishnan, Y. Zheng, and A. X. Liu, "DeepScaling: microservices autoscaling for stable CPU utilization in large scale cloud systems," In *Proceedings of the 13th Symposium on Cloud Computing*, pp. 16-30, 2022. <https://doi.org/10.1145/3542929.3563469>.
- [14] N. Nithiyandam, M. Rajesh, R. Sitharthan, D. Shanmuga Sundar, K. Vengatesan, and K. Madurakavi, K., "Optimization of Performance and Scalability Measures across Cloud Based IoT Applications with Efficient Scheduling Approach," *International Journal of Wireless Information Networks*, vol. 29, no. 4, pp. 442-453, 2022. <https://doi.org/10.1007/s10776-022-00568-5>.
- [15] S. Henning, and W. Hasselbring, "A configurable method for benchmarking scalability of cloud-native applications," *Empirical Software Engineering*, vol. 27, no. 6, pp. 1-42, 2022. <https://doi.org/10.1007/s10664-022-10162-1>.
- [16] S. N. Srirama, M. Adhikari, and S. Paul, "Application deployment using containers with auto-scaling for microservices in cloud environment" *Journal of Network and Computer Applications*, vol.160, pp. 102629-102641, 2022. <https://doi.org/10.1016/j.jnca.2020.102629>.
- [17] A. Al-Said Ahmad, and P. Andras, "Cloud-based software services delivery from the perspective of scalability," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 36, no. 2, pp. 53-68, 2021. <https://doi.org/10.1080/17445760.2019.1617864>.
- [18] C. Li, J. Liu, B. Lu, and Y. Luo, "Cost-aware automatic scaling and workload-aware replica management for edge-cloud environment," *Journal of Network and Computer Applications*, vol. 180, pp. 103017, 2021. <https://doi.org/10.1016/j.jnca.2021.103017>.
- [19] B. Liu, Y. Chen, "A scalable fine-grained analytic model for container cloud data centres," *Int. J. Internet Technol. Secur. Trans.*, vol. 9, no. 4, pp. 355–389, 2019. <https://doi.org/10.1504/IJTST.2019.102794>.
- [20] N. Nithiyandam, M. Rajesh, R. Sitharthan, D. Shanmuga Sundar, K. Vengatesan, and K. Madurakavi, K., "Optimization of Performance and Scalability Measures across Cloud Based IoT Applications with Efficient Scheduling Approach," *International Journal of Wireless Information Networks*, vol. 29, no. 4, pp. 442-453, 2022. <https://doi.org/10.1007/s10776-022-00568-5>.
- [21] M. Imdoukh, I. Ahmad, M. G. Alfaiakawi, "Machine learning-based auto-scaling for containerized applications," *Neural Computing and Applications*, vol. 32, no. 13, pp. 9745-9760, 2020. <https://doi.org/10.1007/s00521-019-04507-z>.
- [22] S. Chouliaras, and S. Sotiriadis, "Auto-scaling containerized cloud applications: A workload-driven approach," *Simulation Modelling Practice and Theory*, vol. 121, pp. 102654, 2022. <https://doi.org/10.1016/j.simpat.2022.102654>.
- [23] A. Wang, S. Chang, H. Tian, H. Wang, H. Yang, H. Li, and Y. Cheng, "{FaaSNet}: Scalable and Fast Provisioning of Custom Serverless Container Runtimes at Alibaba Cloud Function Compute," *USENIX Annual Technical Conference (USENIX ATC 21)*, pp. 443-457, 2021.
- [24] D. Perri, M. Simonetti, S. Tasso, F. Ragni, and O. Gervasi, "Implementing a scalable and elastic computing environment based on cloud containers," *International Conference on Computational Science and Its Applications*, Springer, pp. 676-689, 2021. [https://doi.org/10.1007/978-3-030-86653-2\\_49](https://doi.org/10.1007/978-3-030-86653-2_49).
- [25] J. Y. Choi, M. Cho, and J. S. Kim, "Employing Vertical Elasticity for Efficient Big Data Processing in Container-Based Cloud Environments," *Applied Sciences*, vol. 11, no. 13, pp. 6200, 2021. <https://doi.org/10.3390/app11136200>.
- [26] I. Kabashkin, "Availability of applications in container-based cloud PaaS architecture," *International Conference on Reliability and Statistics in Transportation and Communication*, Springer, pp. 241–248, 2018. [https://doi.org/10.1007/978-3-030-12450-2\\_22](https://doi.org/10.1007/978-3-030-12450-2_22).
- [27] S. El Kafhali, K. Salah, "Performance modeling and analysis of internet of things enabled healthcare monitoring systems," *IET Netw.* vol. 8, no. 1, pp. 48–58, 2019. <https://doi.org/10.1049/iet-net.2018.5067>.
- [28] H. Chen, D. D. Yao, "Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization," vol. 46. Springer, Berlin ,2013. <https://doi.org/10.1007/978-1-4757-5301-1>.
- [29] R. Nelson, "Probability, Stochastic Processes, and Queueing Theory: the Mathematics of Computer Performance Modeling," Springer, Berlin 2013. <https://doi.org/10.1007/978-1-4757-2426-4>.
- [30] K. Salah, S. El Kafhali, "Performance modeling and analysis of hypoeponential network servers," *J. Telecommun. Syst.*, vol. 65, no. 4, pp. 717–728, 2017. <https://doi.org/10.1007/s11235-016-0262-3>.
- [31] Burke, "P.J.: The output of a queuing system," *Oper. Res.*, vol. 4, no. 6, pp. 699–704, 1956. <https://doi.org/10.1287/opre.4.6.699>.
- [32] S. El Kafhal, K. Salah, S. Ben Alla, "Performance evaluation of IoT-fog-cloud deployment for healthcare services," *International Conference on Cloud Computing Technologies and Applications (CloudTech'18)*, IEEE, pp. 1–6 ,2018. <https://doi.org/10.1109/CloudTech.2018.8713355>.
- [33] U. N. Bhat, "An Introduction to Queueing Theory: Modeling and Analysis in Applications, Springer, New York ,2015. <https://doi.org/10.1007/978-0-8176-4725-4>.

# An Improved Breast Cancer Classification Method Using an Enhanced AdaBoost Classifier

Yousef K. Qawqzeh<sup>1\*</sup>, Abdullah Alourani<sup>2</sup>, Sameh Ghwanmeh<sup>3</sup>

Information Technology College, University of Fujairah, Fujairah, 125212, United Arab Emirates<sup>1,3</sup>

Department of Computer Science and Information, College of Science in Zulfi, Majmaah University  
Al-Majmaah, 11952, Saudi Arabia<sup>2</sup>

**Abstract**—The goal of this research is to create a machine learning (ML) classifier that can improve breast cancer (BC) diagnosis and prediction. The principle components analysis (PCA) technique is used in this work to minimize the dimensions of the BC dataset and achieve better classification metrics. The developed classifier outperformed others in terms of F1 score and accuracy score. Using the original BC dataset, four different classifiers are applied to determine the best classifier in terms of performance metrics. The used classifiers were RandomForest, DecisionTree, AdaBoost, and GradientBoosting. The RandomForest classifier obtained (95.7%) f1 score and (94.5%) accuracy score, the DecisionTree classifier obtained (93%) f1 score and (91%) accuracy score, the GradientBoosting classifier obtained (95%) f1 score and (93.5%) accuracy score, and the AdaBoost classifier obtained (95.8%) f1 score and (94.5%). The AdaBoost classifier was utilized to create the final model using the reduced PCA dataset because it scored the highest performance metrics. The developed classifier is named as “pcaAdaBoost”. The optimized pcaAdaBoost achieved higher performance metrics in terms of f1 score (99%) and accuracy score (98.8%). The results reveal that the optimized pcaAdaBoost scored highest performance measures in terms of cross-validation and testing outcomes, with an overall accuracy of (99%). The improved results justify the use of dimensionality reduction in high-dimension datasets to reduce complexity and improve performance measures.

**Keywords**—Breast cancer; diagnosis; prediction; AdaBoost, RandomForest; PCA

## I. INTRODUCTION

Several studies and technologies have been conducted worldwide to screen for and investigate the risk of Breast Cancer (BC). Despite significant advances in screening and patient management, BC represents one of the common malignancies in women worldwide and it was ranked as the second most likely cause of cancer mortality. Statistically, there were 268,600 new cases of BC diagnosed in American women in 2019, with 41,760 deaths [1-3]. BC is a very diverse illness with numerous forms and subtypes. Approximately 95 % of BCs responded to endocrine and targeted therapy, and their prognosis and survival rates are generally favorable. However, the widely used screening instrument is a two-dimensional mammography, which can detect tumors that are too small to perceive. The breast is compressed between two rigid plates in a conventional mammogram, and X-rays have been used to capture images of the breast tissue. Such techniques are invasive, costly, and tedious to conduct. With the advent of new computational power in terms of big data, machine

learning (ML), and data science (DS), scholars have attempted to apply such new computation techniques to the analysis of BC datasets, as well as to develop new promising low cost and fast BC classification techniques. However, to reduce processing time and to increase prediction performance, data reduction techniques are used. Removing irrelevant input data and get rid of redundant inputs would probably enhance classifier's capability in terms of performance measures. In this study, the PCA technique is used in this work to minimize the dimensions of the BC dataset and achieve better classification metrics. The generated classifier outperformed others in terms of F1 score and accuracy score. Using the original BC dataset, four different classifiers are applied to determine the best classifier in terms of performance metrics. Therefore, the following classifiers, RandomForest, DecisionTree, AdaBoost, and GradientBoosting, are used and evaluated. This work utilized the PCA method to reduce features from the original BC dataset. This method improved the performance of the ML model on hand and enabled better data visualization. In this way, the PCA is used to reduce the dimensions of the BC dataset, making it less sparse and more statistically significant.

## II. LITERATURE REVIEWS

Scholars defined ML as a subset of artificial intelligence (AI). It denotes a mathematical model that is used to make decisions or predictions using a training dataset. It is frequently referred to as an evolving prediction model that will improve classification capabilities in a variety of fields including disease diagnosis and screening as in medical industry [4]. It is necessity to reduce the danger of diseases, infections, disorders, or pandemics using a proactive ML model [5-6]. To deal with the increasing complexity of the vast data and convert it into meaningful scientific knowledge for the benefit of humanity, new bioinformatics methods must be developed [7-8]. The use of ML classification techniques in medical diagnosis applications is highly valued [6]. However, traditional classification may not have performed as well as planned, raising the necessity of such investigations that could improve the current classification technologies in medical sectors. The goal of medical AI and ML research is to create applications that use AI technologies to aid practitioners in providing treatment based on better decision making [9-10]. Some research has been conducted on comprehensible AI in order to solve the downfalls of AI analysis tools being black boxes. In comparison to AI systems such as deep learning, XAI can present model's explanations and decision-making

\*Corresponding Author.

capabilities [11]. Many traditional ML approaches for classification problems are used like logistic regression (LR), support vector machine (SVM), decision tree (DT), and RandomForest (RF). The RF is a ML classification method that comprised of several decision trees in an ensemble. The outcome of these DT elections symbolizes the RF decision. Regardless of the used ML algorithm, several evaluation methods such as F1\_score, accuracy\_score, precision, recall, AUC, and ROC have been commonly used to assess the effectiveness of each proposed method [12]. One of the most important areas of medical based ML applications is BC classification. BC has now surpassed lung cancer as being the most prevalent malignancy afflicted in women worldwide [13]. For the prognosis and diagnosis of BC disease, researchers developed a SVMs based classifier in contrast to Bayesian classifiers and ANN. However, they gave implementation summary for the findings of the evaluated classifiers [14]. Another study [15] used the BC dataset to classify BC disease using several ML methods: Knn, RF, SVM, DT, and LR. They evaluate the results of each method and concluded that the SVM algorithm outperformed the others with a performance accuracy of (97.2%). Another study [16] tried to examine the findings and to analyze several ML approaches for the detecting of BC using the same dataset. Another study [17] proposed an efficient recursive neural network (RNN) approach for BC classification using RNN and “Keras-Tuner” enhancement method in which they claimed that, the developed model achieved high performance accuracy. However, the study in [16] showed that Logistic regression classifier beats the other classifiers in predicting BC disease using BC Wisconsin (Diagnostic) data set (BCWD).

### III. RESEARCH METHODS

The current study attempts to minimize the dimensionality

of the dataset before selecting the optimal features to be fed to the classifier. The proposed method employs the simplest model that meets the performance requirements of the complicated models. Accordingly, the dataset shall be dimensionally reduced in order to improve model performance and eliminate extraneous features. Therefore, the proposed model begins with data preprocessing, followed by feature selection, dimensionality reduction, and classification (Fig. 1). In this methodology, four different supervised classification algorithms are used, RandomForest, AdaBoost, GradientBoost, and DecisionTree respectively. For this study, the classifier that obtained the best performance measure (Accuracy\_score, and F1\_score) is selected to perform classification process. However, after selecting the best classifier, the principle components analysis (PCA) procedure is taken place to perform dimensionality reduction. The reduced dataset is then fed to the chosen classifier to implement BC classification. The developed model is then validated using k-fold cross validation, tested using a subset of the original dataset (30%), and lastly, performance measures are evaluated.

#### A. Dataset Description

In this study, BCWD data is being utilized for model development purposes. The data include 31 features along with the class feature (target). The cell nuclei detected in the breast image clip are represented by the independent indices. Moreover, the dependent index contains binary outcome: zero indicates benign, and one describes malignant. The output will be classified as being benign or malignant. However, the shape of the used dataset is (569,31), and its descriptive analysis is shown in Fig. 2. Additionally, the bar chart (Fig. 3) depicts the count of the target variable to be malignant (M) or benign (B).

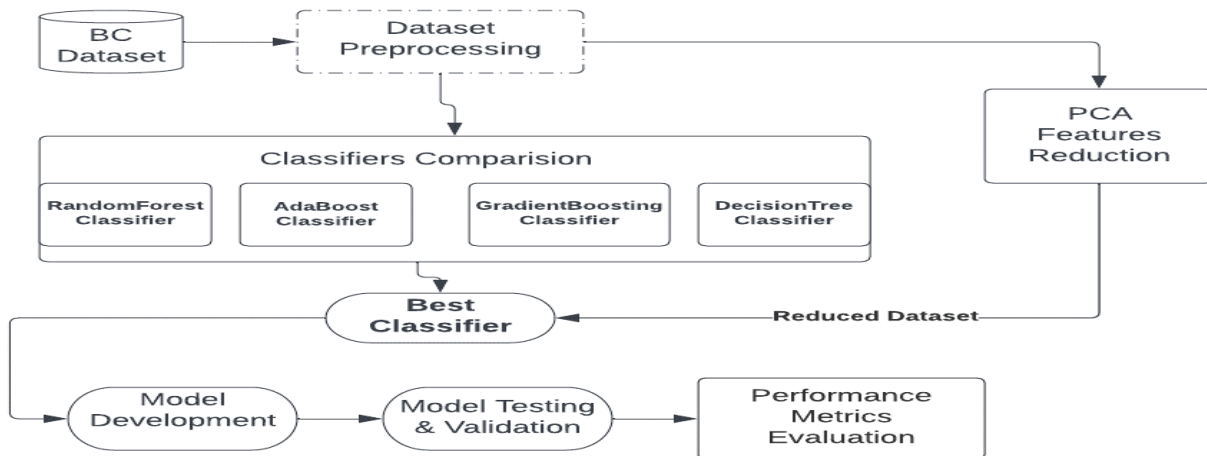


Fig. 1. An enhanced methodology for BC classification using PCA.

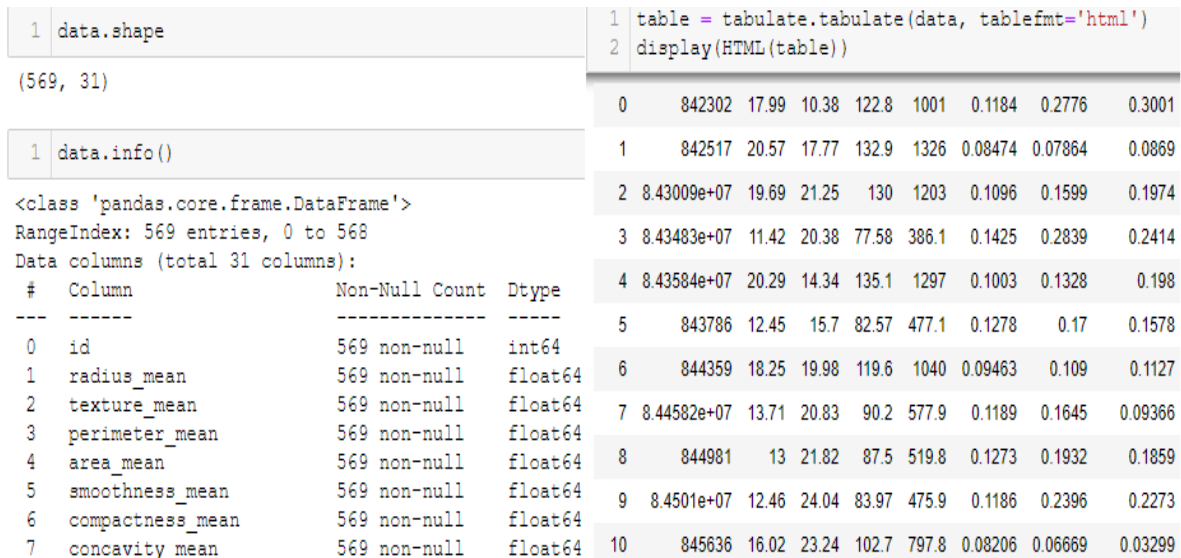


Fig. 2. Descriptive analysis of BC dataset.

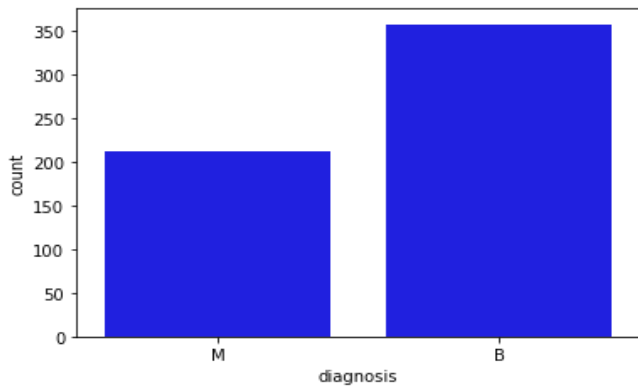


Fig. 3. Count for class variable (diagnosis).

### B. Dimensionality Reduction

Features reduction represents a very important preprocessing stage that eliminates redundancy, inconsistent, and unimportant features to optimize learning, classification accuracy, and minimize training cost [18]. One of the its approaches that diminishes computation cost for the learning process is called “Principal Component Analysis (PCA)”. Moreover, features reduction is useful in several realms because it reduces the computational burden as well as other unfavorable characteristics of high-dimensional areas. Many scholars recommended the utilization of features reduction techniques to improve computational power and to enhance performance accuracy [5]. Therefore, the literature has several uses of dimensionality reduction techniques such as, Zhao and Du [19] in which they advocated for the use of the “feature\_based” spectral-spatial” classification (SSFC) structure. Another study by Xu. Y et. al [20] proposed spontaneous removal of piece picture from side to side deep learning. However, the PCA is commonly used method for features reduction.

### C. Classification

Classification represents the most important task in supervised learning techniques [21-23]. It normally utilized to separate the dataset into a unique class as per the values in the dependent variable [24-25]. To select the best classifier that provides the optimal performance metrics based on BC dataset, four different classifiers are used, namely, RandomForest, Adaboost, Gradientboost, and DecisionTree are used. A brief detail-on each classifier is shown below:

1) *Random forest classifier*: Because of its ease of implementation and high versatility, it is one of the most often used supervised learning algorithms. It is a collection of prediction trees capable of handling large datasets.

2) *Adaboost classifier*: AdaBoost was among the first applications to employ the boosting technology. It accomplishes this by integrating numerous weak classifiers into a single strong classification method.

3) *Gradientboost classifier*: It is an ensemble, functional gradient iterative approach that reduces a “loss function” by repeatedly selecting a function who points towards the negative slope.

4) *Decision tree classifier*: The decision tree can be defined as a supervised learning technique in which it is commonly implemented for solving binary classification problems. However, such a technique bases its decision on some rules.

## IV. RESULTS

In this study, initially four different classifiers were employed to get performance metrics using BC dataset. The established approaches were assessed using accuracy score and F1 score metrics. However, the model with the best metrics was used to develop the enhanced model. After data preprocessing and visualization, the dataset is reduced into two main components namely, first principal component and second principal component. Fig. 4 illustrates the reduced PCA components.



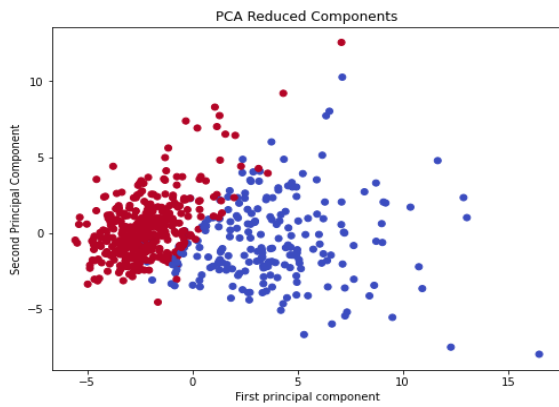


Fig. 4. The reduced PCA components (first & second principal components).

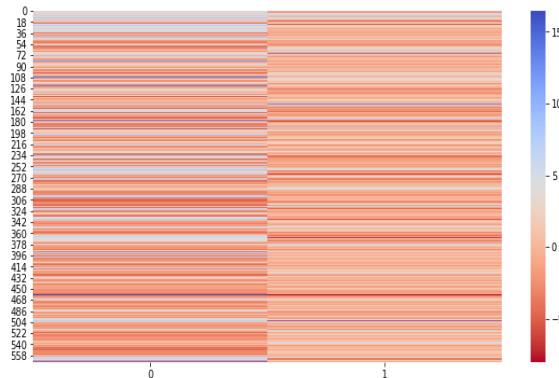


Fig. 5. Heatmap of the reduced PCA features.

As it can be seen in Fig. 4, the reduced dataset consists of two principal components in which they can represent the original dataset with no loss of information. It can be noticed that the reduced dataset can be used to clearly discriminate between the two classes of the target variable ('diagnosis'). The reduced dataset represents a dumpy matrix array in its rows represents the principal components and each column in it relates back to the original indices. Such relationship can be visualized as heatmap (Fig. 5).

The classifier that obtained the higher performance metrics is selected to implement the final model using the reduced PCA dataset. The RandomForest classifier obtained (95.7%) f1\_score and (94.5%) accuracy\_score in which it occupied position number two among the selected classifiers. The DecisionTree classifier obtained (93%) f1\_score and (91%) accuracy\_score. The GradientBoosting classifier obtained (95%) f1\_score and (93.5%) accuracy\_score. Eventually, the AdaBoost classifier obtained (95.8%) f1\_score and (94.5%) accuracy\_score (Table I).

TABLE I. CLASSIFIERS PERFORMANCE METRICS

Classifier	F1-Score	Accuracy-Score
Random Forest	95.7%	94.5%
Decision Tree	93%	91%
Ada-Boost	95.8%	94.5%
Gradient-Boosting	95%	93.5%

## V. DISCUSSIONS

The obtained results showed that AdaBoost classifier outperforms other used classifiers in terms of f1\_score and accuracy\_score. Thereby, it was used to implement the final enhanced model. On other hand, the RandomForest classifier achieved a very similar performance metrics to the AdaBoost classifier in terms of accuracy\_score, but the AdaBoost classifiers achieved higher F1\_score. Hence, the reduced PCA components were fed again to AdaBoost classifier to validate the enhancement made by the reduced dataset features using PCA components. Fig. 6 showed the enhanced performance metrics in which it achieved an overall accuracy of (99%). The new enhanced model is used to make predictions on a new dataset to validate its performance. The developed classifier was able to correctly classify the new data into its correct classes where they were 'Malignant' or 'Benign'. The final accuracy\_score and F1\_score is depicted in Fig. 7. The model obtained a noticeable higher f1\_score (99%) and noticeable higher accuracy\_score (98.8%).

```
1 flower_type = {0:'Malignant', 1:'Benign'}
2
3 flower_index = 7
4
5 y_test_np = np.array(y_test)
6
7 print(f'Actual --> {flower_type[y_test_np[flower_index]]} -- Prediction --> {flower_type[pred[flower_index]]}')

Actual --> Malignant -- Prediction --> Malignant

1 print('Enhanced classification metrics using PCA and AdaBoost Classifier')
2 print('Accuracy_Score is: ')
3 print(accuracy_score(y_test,pred))
4 print('F1_Score is: ')
5 print(f1_score(y_test,pred))

Enhanced classification metrics using PCA and AdaBoost Classifier
Accuracy_Score is:
0.9883040935672515
F1_Score is:
0.9908256880733944
```

Fig. 6. The enhanced BC classification performance metrics.

```
1 print(confusion_matrix(y_test,pred))  
[[ 61  0]  
 [  2 108]]  
  
1 print('AdaBoost Classification Using PCA')  
2 print(classification_report(y_test,pred))  
  
AdaBoost Classification Using PCA  
              precision    recall  f1-score   support  
  
   0           0.97       1.00       0.98         61  
   1           1.00       0.98       0.99        110  
  
 accuracy          0.99         0.99         0.99         171  
 macro avg          0.98         0.99         0.99         171  
 weighted avg          0.99         0.99         0.99         171
```

Fig. 7. The improved performance metrics using AdaBoost and PCA.

As shown in Fig. 7, the developed model obtained higher accuracy score and higher F1 score. Therefore, it can be concluded that, the reduced dataset using PCA components analysis can enhance classification performance in high-dimensions datasets. Furthermore, dimensionality reduction simplifies the classification process in ML, resulting in a better fit to the constructed classifier.

## VI. CONCLUSIONS

This research utilized PCA technique to minimize the input features in the BC dataset seeking better enhancement of BC classification in terms of F1\_score and accuracy\_score. The developed model started with a performance metrics comparison between four supervised classification techniques namely, RandomForest, DecisionTree, AdaBoost, and GradientBoosting. The RandomForest classifier showed (95.7%) f1\_score and (94.5%) accuracy\_score, DecisionTree classifier obtained (93%) f1\_score and (91%) accuracy\_score, GradientBoosting classifier obtained (95%) f1\_score and (93.5%) accuracy\_score, and finally, AdaBoost classifier obtained (95.8%) f1\_score and (94.5%) accuracy\_score. Since the AdaBoost classifier scored the highest performance metrics, it used to implement the final model using the reduced PCA dataset. The developed classifier is named "pcaAdaBoost". The optimized pcaAdaBoost achieved higher performance metrics in terms of F1\_Score (99%) and accuracy\_score (98.8%). The results show that the optimized pcaAdaBoost has delivered the best results in terms of cross-validation and testing. with an overall accuracy of (99%). However, as per future works, the developed classifier should be trained and tested using different datasets to validate its ability to enhance performance metrics. Finally, the developed model is hoped to introduce a predictive tool for early diagnosis and classification of BC in our large society.

## VII. DATA AVAILABILITY

The used data in the development of this model and that is used to support the findings of this research can be accessed online at:

UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set.

## VIII. CONFLICTS OF INTEREST

The authors stated that they do have no potential conflicts to disclose in relation to this study.

## REFERENCES

- [1] R.L Siegel, K.D Miller, and A. Jemal, "Cancer Statistics, 2019," CA Cancer J. Clin, vol. 69, pp 7-34, 2019. doi: 10.3322/caac.21551.
- [2] Y. Qawqzeh and K. Abdus Sattar, "Online Diagnostic Expert System for Detection of Breast Cancer in Saudi Arabia," International Journal of Computer Applications. 113. pp40-47, 2015. 10.5120/19833-1686.
- [3] J.A Basurto-Hurtado, I.A Cruz-Albarran, M. Toledano-Ayala, M.A Ibarra-Manzano, L.A Morales-Hernandez, C.A Perez-Ramirez, "Diagnostic Strategies for Breast Cancer Detection: From Image Generation to Classification Strategies Using Artificial Intelligence Algorithms," Cancers 2022, 14, 3442. <https://doi.org/10.3390/cancers14143442>.
- [4] X-D. Zhang, "Evolutionary computation. In: A matrix algebra approach to artificial intelligence," Springer, Singapore, pp 681-803, 2020.
- [5] A.K.S. Ong et al." Utilization of random forest and deep learning neural network for predicting factors affecting perceived usability of a covid-19 contact tracing mobile application in Thailand," Int. J. Environ. Res. Public Health, 19, 6111, 2022, doi.org/10.3390/ijerph19106111.
- [6] Y. Qawqzeh, M. T. Alharbi, A. Jaradat and K. N. Abdul Sattar, "A review of swarm intelligence algorithms deployment for scheduling and optimization in cloud computing environments," PeerJ Computer Science, vol. 7, pp. e696, 2021.
- [7] D. Rajkumar, S. P. Raja, and A. Suruliandi, "Users' click and book-mark based personalization using modified agglomerative clustering for web search engine," International Journal on Artificial Intelligence Tools, 26, no. 06, 2017, doi: 10.1142/S0218213017300022.
- [8] P. Divya, M. Pavithra, S. Jayalakshmi, P. Praveen kumar, "Application of Random Forest Algorithm in Bio Informatics," International Journal of Information Technology Insights & Transformations, Vol. 5, I 1, 2021, pp16-24.
- [9] M. Alloghani, D. Al-Jumeily, A.J. Aljaaf, M. Khalaf, J. Mustafina, and S.Y. Tan, "The Application of Artificial Intelligence Technology in Healthcare: A Systematic Review," Commun. Comput. Inf. Sci. 2020, 1174, 248-261.
- [10] M. Strzelecki and P. Badura, "Machine Learning for Biomedical Application," Applied Sciences, 2022; 12(4):2022. doi.org/10.3390/app12042022.
- [11] Y. Zhang, Y. Weng, and J. Lund, "Applications of Explainable Artificial Intelligence in Diagnosis and Surgery," Diagnostics 2022, 12, 237. doi.org/10.3390/diagnostics12020237.
- [12] Y.K. Qawqzeh, M.M. Otoom, F. Al-Fayez, I. Almarashdeh, M. Alsmadi and G. Jaradat, "A Proposed Decision Tree Classifier for Atherosclerosis Prediction and Classification," IJCSNS, v 19, issue 12, 2019, pp197-202.
- [13] WHO, "Preventing cancer," Accessed on: Oct. 18, 2022. [online] Available: <https://www.who.int/activities/preventing-cancer>.
- [14] I. Maglogiannis, E. Zafiroopoulos, and I. Anagnostopoulos, "An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers," Applied Intelligence, vol. 30, pp. 24-36, 2009.
- [15] M. Naji and S. El Filali, "Machine learning algorithms for breast cancer prediction and diagnosis," Procedia computer science, 191(2021), pp487-492.
- [16] M. Khan et al., "Machine Learning Based Comparative Analysis for Breast Cancer Prediction," Journal of Healthcare Engineering, Hindawi. V2022, pp1-14. doi.org/10.1155/2022/4365855.
- [17] H. Saleh, S.F. Abd-el ghany, H. Alyami, and W. Alosaimi, "Predicting Breast Cancer Based on Optimized Deep Learning Approach," Hindawi, Computational Intelligence and Neuroscience, Vol 2022, pp1-11. <https://doi.org/10.1155/2022/1820777>.
- [18] S. Velliangiri, S. Alagumuthukrishnan, S.I.T. Joseph, " A Review of Dimensionality Reduction Techniques for Efficient Computation," Procedia Computer Science, V165, pp104-111. ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2020.01.079>.
- [19] W. Zhao and S. Du, "Spectral-Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach," IEEE Trans. Geosci. Remote Sens. 54(8): 2016, 4544-4554.

- [20] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, and E. I. Chang, "Deep Learning of Feature Representation with Multiple Instance Learning for Medical Image Analysis,". State Key Laboratory of Software Development Environment, Key Laboratory of Biomechanics and Mechanobiology of Ministry of Education, Beihang University M," *Icassp* 1: 2014, 1645–1649.
- [21] I.H. Sarker, "Machine Learning: Algorithms,,". *Real-World Applications and Research Directions*. SN COMPUT. SCI. 2, 160, 2021. doi.org/10.1007/s42979-021-00592-x.
- [22] N. Binsaif, "Application of Machine Learning Models to the Detection of Breast Cancer," *Mobile Information Systems*, Hindawi. V 2022, https://doi.org/10.1155/2022/7340689.
- [23] M. Shanbehzadeh, H. Kazemi-Arpanahi, M.B Ghalibaf, A. Orooji, "Performance evaluation of machine learning for breast cancer diagnosis: A case study," *Informatics in Medicine Unlocked*, V 31, 2022, 101009, https://doi.org/10.1016/j.imu.2022.101009.
- [24] Butt, Umair Muneer et al. "Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications," *Journal of healthcare engineering* vol. 2021 9930985. 29 Sep. 2021, doi:10.1155/2021/9930985.
- [25] M.M Khan, T. Tazin, M.Z Hussain, M. Mostakim, T. Rehman, S. Singh, V. Gupta, O. Alomeir, "Breast Tumor Detection Using Robust and Efficient Machine Learning and Convolutional Neural Network Approaches," *Exploration of Human Cognition using Artificial Intelligence in Healthcare*. V 2022, https://doi.org/10.1155/2022/6333573.

#### AUTHORS' PROFILE



YOUSSEF QAWQZEH received the PhD. degree in systems engineering from UKM University, Bangi, Kuala Lumpur, Malaysia, in 2011, where he is currently working as an associate professor in the college of information technology, Fujairah University. He is currently working on several projects in the fields of machine learning, data science, and bioinformatics. He has several publications in international journal and conferences. His research interest includes the early prediction of cardiovascular diseases using the photoplethysmography technique, the development of computer-aided diagnosis systems for early diagnosis of breast cancer using artificial intelligence and machine learning techniques, and the detection and prediction of high-risk diabetics using machine learning and artificial intelligence techniques.



ABDULLAH ALOURANI is an assistant professor at the Department of Computer Science and Information, Majmaah University, Saudi Arabia. He received his Ph.D. in computer science from the University of Illinois at Chicago, United States, his Master's degree in computer science from DePaul University in Chicago, United States, and his Bachelor's degree in computer science from Qassim University, Saudi Arabia. His current research interests are in the areas of software engineering, security, and artificial intelligence. He is a member of ACM and IEEE.



SAMEH GHWANMEH a full Professor of Computer Science and Engineering, and ICT. Obtained his PhD and MS in Computer Engineering from UK, in 1996 and 1993 respectively, and BS degree in Computer Engineering from Jordan, in 1985. More than 24 years of leadership experience in HE institutions. Was and still involved in many administrative and academic positions: University Chancellor (current), Registrar Director, Dean of Faculty of Information Technology, Director of Accreditation, Associate Dean, Program Chair, Academic Adviser to the Minister of Education and Higher Education in Jordan. The published work exceeds 50 international journal papers, and was an invited speaker in different international conferences and workshops. Received many national and international awards. Was involved in curriculum development for a number of undergraduate and graduate programs. Managed and coordinated several research projects that are funded by regional and international agencies such as European Commission.

# Efficient Multimedia Content Transmission Model for Disaster Management using Delay Tolerant Mobile Adhoc Networks

Sushant Mangasuli, Mahesh Kaluti

Dept. of Computer Science and Engineering, PES College of Engineering, Mandya, India

**Abstract**—Natural and manmade disasters such as earthquakes, floods, unprecedented rainfall, etc. pose several threats to our society. The citizens upload disaster information in the form of multimedia content such as pictures, audio, and videos. Efficient information and communication framework are critical for disaster management. Mobile Adhoc Networks (MANET) have been used effectively for disaster management. However, Disaster management prerequisites following Quality of Service (QoS) requirements such as bandwidth, high delivery ratio, low overhead, and minimal latency; however, the existing data transmission scheme induces high latency and overhead among intermediate devices; In order to meet the QoS requirement of disaster management applications in this paper, High Delivery Efficiency and Low Latency Multimedia Content Transmission (HDELL-MCT) scheme for MANETs is presented. Then, an improved buffer management scheme is presented for meeting disaster management performance and latency prerequisites. The experiment is conducted using ONE Simulator, the outcome shows the HDELL-MCT scheme achieves very good performance considering different QoS metrics such as improving delivery ratio by 38.02%, reducing latency by 7.53% and minimizing hop communication overhead by 65.1% in comparison with existing multimedia content transmission model.

**Keywords**—Buffer management; disaster management; Mobile Adhoc Network; opportunistic routing; Quality of Service

## I. INTRODUCTION

Disaster management is unpredictable as the disasters can be caused by humans or can be a natural calamity. In the disaster, many man-made things are damaged or destroyed such as buildings, cellular towers, power grids, network infrastructure, etc. [1], [2]. Due to these damages, there is no proper communication if a person is stuck in a disaster. Moreover, there is congestion in the communication channel due to the data traffic as the people try to contact their family or friends seeking help. The disaster management teams are deployed to rescue the people, and resources and many people use an online platform to share the condition of that area by posting videos and pictures [3]. Due to this, there is more traffic and congestion in the communication networks [4]. Despite having massive technologies, there are many communication problems which include system overload, system failure, and incompatibility between communication systems used by different agencies. These communication problems can be resolved by providing a redundant

communication channel or by providing a separate communication network for disaster management [5], [6].

In disaster management, the Mobile Ad-hoc Networks (MANETs) can play an important role by providing the person's information when the rescue team cannot find the person who is stuck in a disaster [7], [8]. The MANETs contain various mobile nodes that are deployed randomly which can join or leave the given network when it moves. The mobile nodes can communicate opportunistically [9] with other nodes using the wireless link to share the information. Mobile Ad-hoc Networks can add various new devices rapidly. All the devices which are in the network can move freely in any direction. Some examples of MANETs that are currently being used are in military operations, a small conference room, rescue operations on the battlefield, ad-hoc networks, and emergency operations [9].

During the transmission of the nodes from the source to the destination, many challenges are faced by the MANETs [10], [11], & [12]. Some of the challenges include the delivery ratio, multihop, congestion control, and latency. Thus, designing a network in MANETs is a challenging task as there are many problems in its design. Every node in the network acts like a router that can send the packets (i.e., images, videos, etc.) from the source to the destination in an opportunistic manner [13]. The nodes can be of any device like mobiles, laptops, personal computers, etc. Mobile Ad-hoc Networks can range from small networks to very large dynamic networks for provisioning multimedia applications [14]. The nodes communicate with the other nodes in a multihop manner. When the transmitter sends a packet to a given destination node, an intermediate node is used for communication in the network. Hence, every node plays a vital role in sending the packets from the source to the destination [3], [15], [16]. An extensive survey shows existing data transmission scheme induces memory overhead among intermediate for provisioning disaster management applications; thus, inducing poor delivery ratio and latency. This motivated the research to develop an improved multimedia content delivery model namely HDELL-MCT for provisioning disaster management applications. The HDELL-MCT incorporates the location feature into the actual packet and predicts whether the current device will go toward destined groups or not; further, each packet is composed of time validity within which the data has to be delivered to the destined group. Then, an improved buffer management scheme is presented for

meeting disaster management performance and latency prerequisites. The significance of HDELL-MCT is as follows.

- The HDELL-MCT employs a two-level multimedia data transmission model for MANETs.
- The HDELL-MCT model can meet the delivery ratio and latency requirement of disaster management applications.
- The HDELL-MCT improves delivery ratio, and intermediate hop overhead, and reduces latency in comparison with the existing multimedia data transmission model.

The manuscript is arranged as follows. In section II, various existing data transmission model for wireless and Mobile Adhoc networks is discussed, and also identified research issues for developing an improved multimedia content transmission model. In section III, the proposed high delivery efficiency and low latency multimedia content transmission model have been presented for disaster management using Mobile Adhoc networks. In section IV, the simulation outcome of HDELL-MCT and existing MCT through The One simulator is validated. In the last section, the result is concluded with a future research direction.

## II. LITERATURE SURVEY

Here are various existing data transmission schemes for provisioning disaster management applications for wireless and Mobile Adhoc networks. In [5], a routing protocol, TA-AOMDV has been proposed which can adjust to the movement of the node at high speed to achieve a good quality of service. In this protocol, an algorithm is used which takes the node resources for the selection of the path and also to connect the stability probability between any two nodes. This protocol uses a technique that constantly brings up-to-date routing strategies using the evaluation of the link stability of the nodes. In [6], a routing scheme known as BARS has been presented which can avoid the overcrowding of the bandwidth problems in network paths and make space in the queues to store the information in the cache. The available storage of the cache must be evaluated before the transmission of the data takes place. This model uses a technique of feedback through which it can identify the traffic source to adjust the data rate using the bandwidth and hold the data in a queue so it can be sent to the routing path. In [7], a study on MANET has been evaluated using the AODV protocol, which evaluates the outcome of the route requesting parameters. This protocol was compared with the OLSR protocol using the performance metrics such as energy consumption, latency, and other metrics. The comparison tells that the OLSR model has some latency when compared with the AODV model.

In [8], a routing protocol has been proposed using the delay-tolerant MANET, in which the virtual nodes are carefully chosen based on the delivery to the destination node, only if the node is not delivered using the MANET protocol. The transmission of the node is done from the source node to the destination node. In [10], a routing protocol has been proposed opportunistic network, which generates a data forwarding model in the communication of the data where it

evaluates the rules of transferring data in a partition scheme for node activity. To achieve the trade-off between the transmission efficiency and system overhead a different approach has been used. A free movement degree model is developed depending on which a utility capacity is concluded to pick nodes for transmission of the information [11]. In [12], focused on addressing the congestion at the node and link-level adopting reactive routing mechanism using random early detection and expanding ring searching models to improve packet delivery ratio with minimal latency. In [13], an energy-efficient version of the FCSG, EFCSG has been proposed for the OppNets. This protocol uses the fuzzy controller with fuzzy features to check the nodes that can be considered for the routing process.

In [15], a technique, TBSMR is used to improve the Quality of Service in MANETs performance. This technique uses different factors like malicious detection of nodes, control of congestion, packet loss reduction, etc. to strengthen the Mobile Ad-hoc Networks quality of service. In [3], a method, namely MMDSR (multipath multimedia dynamic source routing) has been modified from the previous version to include the tie strength in the decision which is done using the forwarding algorithm; where the model finds a trade-off between the QoS and the reputation between the users who arrange the path in the MANET. This method increases the reputation metric without affecting the QoS. However, the major drawback of the existing data transmission scheme is that they achieve poor data delivery ratio and induce memory management overhead considering delay-tolerant disaster management application; For overcoming research issues an improved multimedia data transmission scheme for provisioning disaster management through Mobile Adhoc Network.

## III. HIGH DELIVERY EFFICIENCY AND LOW LATENCY MULTIMEDIA CONTENT TRANSMISSION MODEL FOR DISASTER MANAGEMENT USING MOBILE ADHOC NETWORKS

This section presents high delivery efficiency and low latency multimedia content transmission model for disaster management using Mobile Adhoc networks. In traditional MANETs, the messages are generally broadcasted to every neighboring device available leading to high congestion in the network. In addressing this work employ group-casting-based communication where groups represent a location that can provide services; Any MANET device that resides within a group will become a possible receiver. In group-casting, all information's are addressed to a respective group and each MANET device that resides within that group will receive the data packets; thus, it is important to outline groups in MANETs. Further, information should incorporate the group's information; thus, every MANET device that resides within groups will receive packets.

### A. MANET Device Group Classification

Here we define the groups for MANETs considering delay-tolerant prerequisite of disaster management applications. The groups define MANET devices that reside in the same region and have coverage to communicate the data packets. In disaster management applications the groups are represented through two-dimensional geographical coordinates. Here the MANET

devices randomly transmit information to destined groups; thus, aiding in reducing broadcasting overhead by preventing unwanted message transmission. Most of the existing group-casting-based data transmission scheme assumes groups through circular pair, these models generally increase their circular radius for reaching faraway MANET devices; thus, resulting in bandwidth wastage. In a standard group-casting-based data transmission model the groups must be predefined; however, in this work, the message carries coordinates information which is the address of the destined device. In MANETs, the device is highly mobile and is aware of adjacent device locations; however, each device can share information among them through intermediate nodes. Therefore, every information must have information about destined groups. Along with the data transmission scheme must be in a position to communicate information to a destined group. The data transmission scheme should be in a position to establish MANET device is within the group or not in a reliable manner. The step involved in establishing whether a destined MANET device is within-group or not is described in Algorithm 1.

**Algorithm 1. MANET Device Group Classification**

Step 1	<b>For each</b> collected data with respective MANETs, the device <b>do</b>
Step 2	<b>Obtain</b> the MANETs device location $(x, y)$
Step 3	<b>Compute</b> vertical segment of network with $x$ - coordinate of respective MANETs device's location
Step 4	<b>Establish</b> intersected array among the vertical segment and the receiver's group of MANETs.
Step 5	<b>For each</b> intersected region within the array <b>do</b>
Step 6	<b>If</b> $y$ -coordinate of the intersected region is lesser when compared with $y$ -coordinate of the destination's region <b>then</b>
Step 7	<b>Eliminate it</b> within the array
Step 8	<b>End if</b>
Step 9	<b>End for each</b>
Step 10	<b>If</b> the intersected number is not odd, <b>then</b>
Step 11	<b>get</b> 0 (i.e., MANETs device is not the receiver of corresponding data and is outside the group)
Step 12	<b>Else</b>
Step 13	<b>get</b> 1 (i.e., the current MANETs device is the receiver of corresponding data)
Step 14	<b>End if</b>
Step 15	<b>End for each</b>

**B. Multimedia Data Transmission Scheme**

The objective of HDELL-MCT is to effectively convey the group-casting information to all the MANET nodes that reside within the destined group considering a certain session period. As the MANET device are mobile, they frequently dynamically change groups, thus, incorporating session information is very important. As a result, the HDELL-MCT model delivers to the device that can reach the destination group within the stipulated time. On the other side, the existing routing strategy delivers the packet to detained group according to the device ID defined. However, group association is dependent on MANET device coordinates which vary according to their mobility. Thus, every MANET device must obtain session validity and group address information for routing to be successful to deliver to the intended destination. The main challenge of incorporating session validity into HDELL-MCT, here the session validity is obtained based on

message creation time and its maximum time limit i.e., content validity (CV).

The HDELL-MCT model is composed of two-level. In level 1, an effective carry-forward mechanism toward the destined group. In level 2, message delivery to all MANET device that resides within the region/group. The graphical representation of the two-level HDELL-MCT model is shown in Fig. 1.

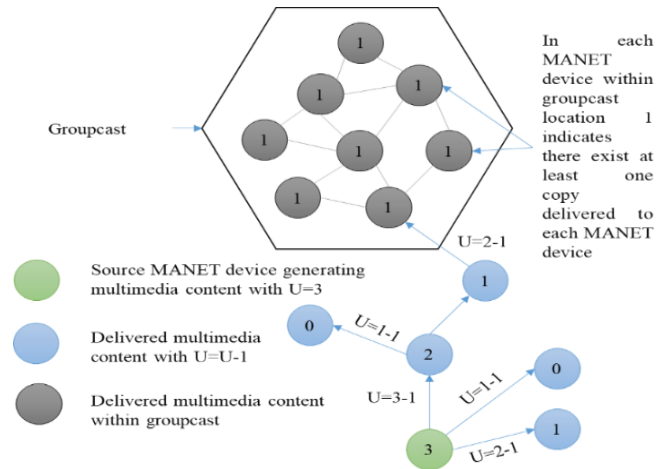


Fig. 1. Graphical representation of HDELL-MDC scheme.

1) *Level 1- Forwarding Multimedia Content toward the destination group:* In this work, a similar multi-copy spray and wait (MCSW) technique [17] is used. The adoption of MCSW aid in reducing communication overhead and improves message delivery; thus, aids in providing an efficient manner of reaching the destination group. Here every time a message is created a token  $U$  is attested to the message. The parameter  $U$  will act as an identifier for representing the size of message replication considered. The parameter  $U$  will be reduced by 1 every time it encounters a MANET device. Once  $U$  reaches to 0, the message will be no longer communicated defining the message as expired, and the message will be deleted from the local memory.

2) *Level 2- Deliver Multimedia Content to all MANET devices within the group:* Here the message is communicated to all the MANET devices within the group by employing an efficient flooding mechanism. The HDELL-MCT model first flood the information to all MANET device and every device keep a copy of the content (as described in line 14-16 and line 21-24 of Algorithm 2). In HDELL-MCT, if a replicated multimedia content goes beyond the destined group, the multimedia content is forwarded back to the MANET device that resides within the group (this step is described in line 20 of HDELL-MCT Algorithm 2); until  $U$  becomes 0 i.e., expires. Finally, the message is removed from the memory (as described in line number 4 of HDELL-MCT Algorithm 2). This aids in reducing memory and network overhead by preventing information exchange beyond the destined group. The two-level HDELL-MCT is described in Algorithm 2. In HDELL-MCT the value of  $U$  can be pre-configured according

to disaster and safety-related application prerequisites such as operating environment, mobility, density, and so on. However, in this work, the U is configured dynamically and device density impacting delivery ratio and latency is a major focus for HDELL-MCT.

**Algorithm 2. Multimedia Data Transmission Scheme**

```

Step 1 For each interaction among two MANET devices
       within the network do
Step 2   One device becomes the sender  $T_1$  and other becomes
       the receiver  $T_2$ 
Step 3 If the MANET device is the sender  $T_1$  then
Step 4   Remove valid multimedia content from the memory
Step 5   Apply memory optimization scheme
Step 6 For each multimedia content within the memory do
Step 7 If multimedia content is already present within  $T_2$ 
       memory
Step 8   Move on to the next multimedia content present within
       the for loop
Step 9 End if
Step 10 If ( $U > 0$ ) then
Step 11  $U \leftarrow U - 1$ 
Step 12 Send a replicated multimedia content to  $T_2$ 
Step 13 Else if ( $U = 0$ ) then
Step 14 If  $T_1$  is within the receiver group (see algorithm 1) then
Step 15 Send a replicated multimedia content to  $T_2$ 
Step 16 End if
Step 17 End if
Step 18 End for each
Step 19 Else if the MANET device is the receiver  $T_1$  then
Step 20 For each received content do
Step 21 If  $T_2$  is within the receiver group (use algorithm 1) then
Step 22 Transmit the multimedia content to the application
       layer
Step 23 Keep a replicated multimedia content in the memory
Step 24  $U \leftarrow 0$ 
Step 25 Else if  $T_2$  is outside the receiver group (use algorithm
       1) then
Step 26 Keep multimedia content in the memory
Step 27  $U \leftarrow U - 1$ 
Step 28 End if
Step 29 End for each
Step 30 End if
Step 31 Change scenario (sender  $T_1$  and receiver  $T_2$ ) and return
       to step 3 at least once
Step 32 End for each

```

**C. Memory Optimization Model**

Algorithm 2 encompasses the memory optimization problem in line 5. The parameter U has been used for delivery efficiency considering limited bandwidth availability, the memory optimization is done using priority optimization. Here we route the packet based on First Come First Serve (FCFS) with Minimal Content Validity (MCV). The HDELL-MCT scheme uses very less resources with a high delivery ratio and less latency when compared with the recent standard data transmission scheme which is experimentally proven next section below.

**IV. SIMULATION RESULT AND ANALYSIS**

Here experiment is conducted for validating the performance of the proposed High Delivery efficiency and low latencies multimedia content transmission (HDELL-MCT)

scheme and standard data transmission [3]. The delivery ratio, intermedia hop forwarding overhead, and latencies by varying network density size are performance metrics considered for validating models. To validate the performance of HDELL-MCT over existing MCT schemes, simulation is conducted using real-time scenarios through The One simulator [18]. This work uses QSMVM (QoS-based Socio-Aware Multi-metric) routing scheme as an existing MCT scheme for comparison because it performed significantly well with better packet delivery and reasonable latency in comparison with various existing baseline data transmission models such as Social aware Content-based Opportunistic Routing Protocol (SCORP) [19], Prophet [20], Epidemic [21], and other data transmission protocol [11].

**A. Simulation Parameter Configuration**

The scenario namely, Helsinki comprised of synthetic traces, is based on the one presented in [3], where nodes are placed on the map of Helsinki city (see Fig. 2). We considered a variable number of MANET devices (i.e., 20, 40, and 80), to assess the performance for different densities. During the simulated 12 h time-period, the MANET device moves on the map roads at an average speed of 50 km/h, between random locations, and with random pause times between 5 and 15 minutes. Each MANET device randomly generates many posts (i.e., to grant the participation of all the MANET devices in the network, every MANET device generates at least one message per day) and generates a message depending on its location at that moment. The parameters taken in this work are based on [3] and Table I shows the main parameters used in the simulated safety-related (disaster management) application scenarios.



Fig. 2. The traffic pattern of devices placed across the map of Helsinki [8].

TABLE I. SIMULATION PARAMETER CONSIDERED

Parameters	Configured
Message Generation Rate	1000 messages per simulation
Message Size	[12] MB
TTL	12 h
Network Interface	IEEE 802.11 a, b, g, n, p
Transmission Range	200 meters

**B. Latency Performance**

Fig. 3 shows the latency induced for transmitting data in the network using QSMVM and HDELL\_MCT scheme considering varied MANET device sizes of 20, 40, and 80. From the result, we can see the HDELL\_MCT model achieves much lesser latency for transmitting packets when compared with the QSMVM scheme. A latency reduction of 7.53% is achieved using HDELL\_MCT over the QSMVM multimedia data transmission scheme.

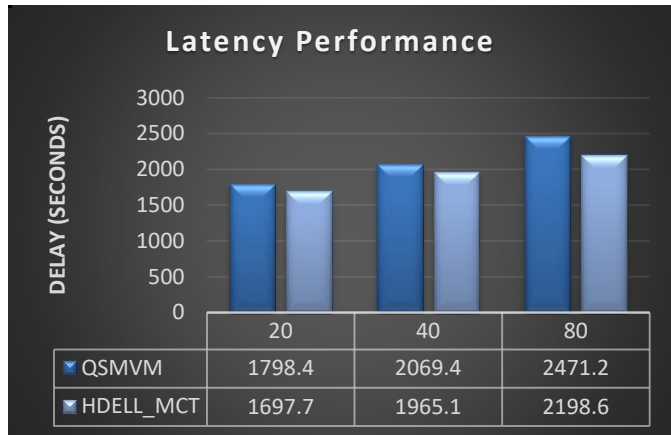


Fig. 3. Multimedia content transmission latency performance for varied MANETs devices.

**C. Delivery Ratio Performance**

Fig. 4 shows the delivery ratio performance for transmitting data in the network using QSMVM and HDELL\_MCT scheme considering varied MANET device sizes of 20, 40, and 80. From the result, we can see the HDELL\_MCT model achieves much superior delivery ratio performance when compared with the QSMVM scheme. A delivery ratio performance improvement of 38.02% is achieved using HDELL\_MCT over the QSMVM multimedia data transmission scheme.

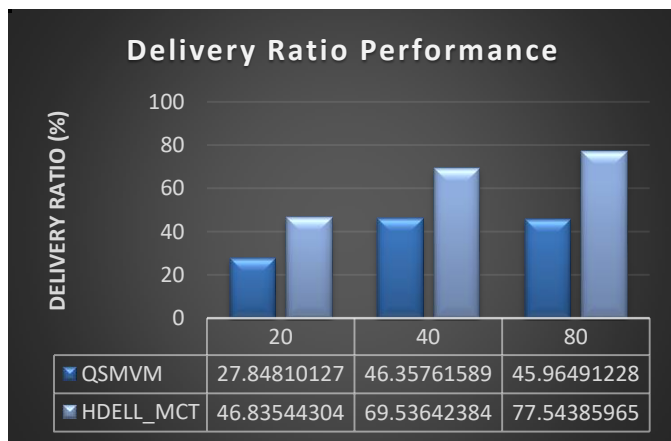


Fig. 4. Multimedia content transmission delivery ratio performance for varied MANETs devices.

**D. Number of Hop-Count Performance**

Fig. 5 shows the hop count performance for transmitting data in the network using QSMVM and HDELL\_MCT scheme considering varied MANET device sizes of 20, 40, and 80. From the result, we can see the HDELL\_MCT model uses less

number of a forwarder for transmitting packets when compared with the QSMVM scheme; thus reducing the control channel computation overhead of the network. A computation overhead reduction of 65.19% is achieved using HDELL\_MCT over the QSMVM multimedia data transmission scheme.

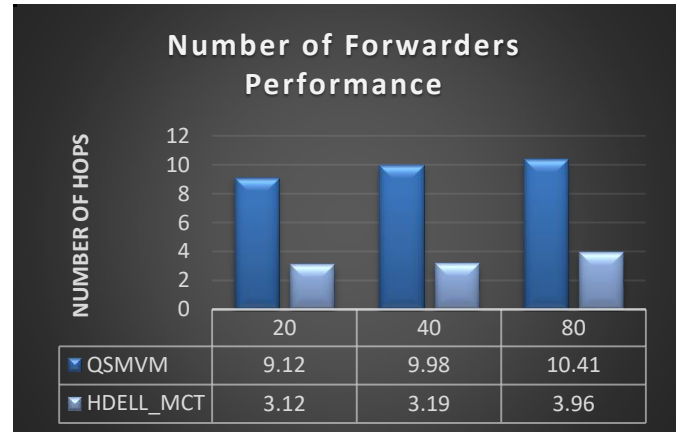


Fig. 5. Multimedia content transmission Hop count performance for varied MANETs devices.

From the overall result attained it can be seen the HDELL\_MCT scheme achieves superior packet delivery ratio performance with fewer latencies and computation overhead under varied density; thus, are suitable for provisioning safety-related (disaster management) applications.

**V. CONCLUSION**

In this paper an effective MCT model namely the HDELL-MCT model is presented for provisioning disaster management applications. The MCT is designed considering improving delivery ratio and reducing latency. In order to meet the QoS requirement of disaster management applications in this paper, High Delivery Efficiency and Low Latency Multimedia Content Transmission (HDELL-MCT) scheme for MANETs is presented. Then, an improved buffer management scheme is presented for meeting disaster management performance and latency prerequisites. The HDELL-MCT employed a two-level multimedia content transmission algorithm for meeting the prerequisite of disaster management applications. Further, HDELL-MCT uses the location-aware feature, group-casting-based efficient flooding, and using priority-based memory optimization scheme balancing delivery ratio and latency reduction performance. The experiment is conducted using ONE Simulator, the outcome shows the HDELL-MCT scheme achieves very good performance considering different QoS metrics such as improving delivery ratio by 38.02%, reducing latency by 7.53% and minimizing hop communication overhead by 65.1% in comparison with existing multimedia content transmission model. Experiment outcome shows the HDELL-MCT model improves delivery ratio, uses less number of intermediate hop MANET devices, and reduces latency in transmitting multimedia data in comparison with the existing data transmission scheme for MANETs.

Future work would consider designing an adaptive propagation model for MCT considering diverse mobility patterns and the presence of obstacles in the line of sight along



with, constructing an effective video encoding technique for improving storage and bandwidth efficiency of delay-tolerant MANETs.

#### ACKNOWLEDGMENT

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript. On behalf of all authors, the corresponding author states that there is no conflict of interest. I would like to thank Dr. Mahesh Kaluti for his valuable and constructive suggestions during all the aspects of our study and for their help in writing the manuscript.

#### REFERENCES

- [1] Manolopoulos, I.; Loukatos, D.; Kontovasilis, K. A Versatile MANET Experimentation Platform and Its Evaluation through Experiments on the Performance of Routing Protocols under Diverse Conditions. *Future Internet*, 14, 154. <https://doi.org/10.3390/fi14050154>, 2022.
- [2] Quy, V.K., Nam, V.H., Linh, D.M. et al. Routing Algorithms for MANET-IoT Networks: A Comprehensive Survey. *Wireless Pers Commun* (2022). <https://doi.org/10.1007/s11277-022-09722-x>.
- [3] Palacios Jara, E.; Mohamad Mezher, A.; Aguilar Igartua, M.; Redondo, R.P.D.; Fernández-Vilas, A. QSMVM: QoS-Aware and Social-Aware Multimetric Routing Protocol for Video-Streaming Services over MANETS. *Sensors* 2021, 21, 901. <https://doi.org/10.3390/s21030901>.
- [4] Mahiddin, N.A., Sarkar, N.I. & Cusack, B. An Internet Access Solution: MANET Routing and a Gateway Selection Approach for Disaster Scenarios. *Rev Socionetwork Strat* 11, 47–64. <https://doi.org/10.1007/s12626-017-0004-3>, 2017.
- [5] Zheng Chen, Wenli Zhou, Shuo Wu and Li Cheng, "An Adaptive on-Demand Multipath Routing Protocol With QoS Support for High-Speed MANET," *IEEE Access*, Vol. 8, pp:44760-44773, doi: 10.1109/ACCESS.2020.2978582, 2020.
- [6] Nousheen Akhtar, Muazzam A. Khan, Ata Ullah, Muhammad Younus Javed, "Congestion Avoidance for Smart Devices by Caching Information in MANETS and IoT," *IEEE Access*, Vol. 7, pp: 71459-71471, doi: 10.1109/ACCESS.2019.2918990, 2019.
- [7] Priyambodo, T.K.; Wijayanto, D.; Gitakarma, M.S. Performance Optimization of MANET Networks through Routing Protocol Analysis. *Computers*, 10, 2. <https://doi.org/10.3390/computers10010002>, 2021.
- [8] Kang, M.W.; Chung, Y.W. An Improved Hybrid Routing Protocol Combining MANET and DTN. *Electronics*, 9, 439. <https://doi.org/10.3390/electronics9030439>, 2020.
- [9] R. Dalal, M. Khari, J. P. Anzola and V. García-Díaz, "Proliferation of Opportunistic Routing: A Systematic Review," in *IEEE Access*, vol. 10, pp. 5855-5883, 2022, doi: 10.1109/ACCESS.2021.3136927.
- [10] Qifei Zhang, Ying Song, Baolin Sun, Zhifeng Dai, "Design of Routing Protocol for Opportunistic Network Based on Adaptive Motion," *IEEE Access*, Vol. 8, pp:18228-18239, doi: 10.1109/ACCESS.2020.2968598, 2020.
- [11] J. Dede, A. F'orster, E. Hern'andez-Orallo, J. Herrera-Tapia, K. Kuladinithi, V. Kuppusamy, P. Manzoni, A. bin Muslim, A. Udugama, and Z. Vatasdas, "Simulating opportunistic networks: Survey and future directions," *IEEE Communications Surveys Tutorials*, vol. 20, no. 2, pp. 1547–1573, Secondquarter 2018.
- [12] Durr-e-Nayab, M. H. Zafar and A. Altalbe, "Prediction of Scenarios for Routing in MANETs Based on Expanding Ring Search and Random Early Detection Parameters Using Machine Learning Techniques," in *IEEE Access*, vol. 9, pp. 47033-47047, 2021, doi: 10.1109/ACCESS.2021.3067816.
- [13] Khuram Khalid, Isaac Woungang, Sanjay Kumar Dhurandher, Jagdeep Singh, "Energy-Efficient Fuzzy Geocast Routing Protocol for Opportunistic Networks," *Advanced Information Networking and Applications, Proceedings of the 35th International Conference on Advanced Information Networking and Applications (AINA-2021), Volume 1*, pp:553-565, doi: 10.1007/978-3-030-75100-5\_48, 2021.
- [14] Badhusha SM, Geetha BG, Prabhu P, Vasanthi R. Improved video streaming using MSVC and nonoverlapping zone routing multipath propagation over MANETs. *Int J Commun Syst*. 2018;e3578. <https://doi.org/10.1002/dac.3578>.
- [15] Mohammad Sirajuddin, Ch. Rupa, Celestine Iwendi, Cresantus Biamba, "TBSMR: A Trust-Based Secure Multipath Routing Protocol for Enhancing the QoS of the Mobile Ad Hoc Network," *Hindwai, Security and Communication Networks*, pp:1-9, doi: 10.1155/2021/5521713, 2021.
- [16] Walunjkar, Gajanan & Anne, Koteswara. (2020). Performance analysis of routing protocols in MANET. *Indonesian Journal of Electrical Engineering and Computer Science*. 17. 1047. [10.11591/ijeecs.v17.i2.pp1047-1052](https://doi.org/10.11591/ijeecs.v17.i2.pp1047-1052).
- [17] Zhang L., Yu C., Jin H. (2012) Dynamic Spray and Wait Routing Protocol for Delay Tolerant Networks. In: Park J.J., Zomaya A., Yeo SS., Sahni S. (eds) *Network and Parallel Computing. NPC 2012. Lecture Notes in Computer Science*, vol 7513. Springer, Berlin, Heidelberg.
- [18] B. Zanj, M. Belegu and A. Rista, "Simulating the mobile network behavior by applying DTN in ONE simulator of the use case of a real town communication," 2019 8th International Conference on Modern Power Systems (MPS), Cluj-Napoca, Cluj, Romania, 2019, pp. 1-9, doi: 10.1109/MPS.2019.8759750.
- [19] W. Moreira, P. Mendes, and S. Sargento, "Social-aware opportunistic routing protocol based on users interactions and interests," *Ad Hoc Networks*, vol. 129, pp. 100–115, 2014.
- [20] R. Lent, "Performance Evaluation of the Probabilistic Optimal Routing in Delay Tolerant Networks," *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, Dublin, Ireland, 2020, pp. 1-6, doi: 10.1109/ICC40277.2020.9149316.
- [21] Halikul Lenando, Mohamad Alrfaay, "EpSoc: Social-Based Epidemic-Based Routing Protocol in Opportunistic Mobile Social Network", *Mobile Information Systems*, vol. 2018, Article ID 6462826, 8 pages, 2018. <https://doi.org/10.1155/2018/6462826>.

# The Effect of Thermal and Electrical Conductivities on the Ablation Volume during Radiofrequency Ablation Process

Mohammed S. Ahmed<sup>1</sup>, Mohamed Tarek El-Wakad<sup>2</sup>, Mohammed A. Hassan<sup>3</sup>

Department of Biomedical Engineering-Faculty of Engineering, Helwan University, Cairo, Egypt<sup>1</sup>

Department of Biomedical Engineering-Faculty of Engineering & Technology, Future University in Egypt, Cairo, Egypt<sup>2</sup>

Department of Biomedical Engineering-Faculty of Engineering, Helwan University, Cairo, Egypt<sup>3</sup>

**Abstract**—Radiofrequency ablation (RFA) is the treatment of choice for certain types of cancers, especially liver cancer. However, the main issue with RFA is that the larger the tumor volume, the longer the ablation period. That causes more pain for the patient, so the surgeons perform a larger number of ablation sessions or surgeries. The current commonly used electrode material, nickel-titanium alloy, used in RFA is characterized by low thermal and electrical conductivities. Using an electrode material with higher electrical conductivity and thermal conductivity provides more thermal energy to tumors. In this paper, we design two models: a cool-tip RF electrode and a multi-hook RF electrode, which aim to study the effect of the thermal and electrical conductivities of the electrode material on ablation volume. Gold, silver, and platinum have higher thermal and electrical conductivity than nickel and titanium alloy, and therefore we studied the effect of these materials on the ablation volume using two different designs, which are the RF cooling tip electrode and the multi-hook electrode. The proposed model reduces the ablation time and damages healthy tissue while increasing the ablation volume with values ranging from 2.6 cm<sup>3</sup> to 15.4 cm<sup>3</sup>. The results show ablation volume increasing with materials characterized by higher thermal and electrical conductivities and thus reducing patient pain.

**Keywords**—Radiofrequency ablation (RFA); finite element method (FEM); COMSOL; Cool-tip RF electrode; multi-hooks electrode; large tumor ablation

## I. INTRODUCTION

Cancer is one of the leading causes of death in both men and women, with 11.3 million deaths in 2020 [1]. The second most common cause of death in Egypt [2]. Liver cancer is the fourth most common type of cancer worldwide [3] and the second most common cancer in Egypt [2]. The most common way to treat liver tumors is resection (RES), microwave ablation (MWA), and radiofrequency ablation (RFA) [4]. RFA is based on passing high-frequency (400-500 kHz) electric currents through the tissue between two excitation electrodes, causing the tissues to heat up, which results in tissue ablation [5, 6]. The main limitation with RFA is that as tumor volume increases the whole-tumor ablation duration increases, which the patient's pain. Therefore, RFA is effective for small tumors (i.e., tumors with a diameter of 3 cm), but still not very effective for large tumors [7, 8]. Radiofrequency (RF) electrode materials should be biocompatible with the human body and have low relative permittivity to be compatible with

any device [9, 10]. Fang et al., suggested several factors contribute to increased ablation volume with reduced ablation time, such as electrode design and the material of the electrode [11]. Electrodes made of materials with high thermal conductivity and high electrical conductivity can contribute to increased ablation volume and reduced time of ablation [12, 13]. Current electrode materials used in RFA are characterized by low thermal and electrical conductivity, such as nickel-titanium alloy [14].

The finite element model (FEM) is a numerical technique used to perform finite element analysis (FEA) of any given physical phenomenon [15]. Alemayehu, et al., have simulated a medical problem about bones and joints using FEM to solve the problems of cartilage erosion [16]. In [17], the authors have used FEM in various fields of dentistry for calculating the strength and behavior of structures and have succeeded in calculating deflection, stress, vibration, and buckling behavior. While on the other side, the authors simulated FEM to ablate breast cancer by using RFA and succeeded in abating 80% of the tumor [18]. In [19], the authors designed FEM to ablate liver tumors by using cool-tip RF electrodes.

Many studies are conducted in ex-vivo studies [20, 21] and clinical studies [22, 23] on the effect of electrode materials on volume ablation due to their electrical and thermal properties. Thermal conductivity gets high attention for researchers in theory because, as the thermal conductivity of the material increases, so the thermal emission capacity of the material does [24, 25]. Several theoretical studies compared materials that have high thermal conductivity with those of low thermal conductivity [26, 27], while other several studies compared materials that have high electrical conductivity with those that have low electrical conductivity [28-30]. These studies found that the use of materials of high thermal and electrical conductivities helps to increase the ablation volume while reducing the ablation time.

The aim of this paper is to study the effect of different materials having high thermal conductivity and high electrical conductivity for RF electrodes on the ablation volume using a FEM. Although our results are specific to the case of RFA of the liver, they may also be important for other ablation applications, such as kidney cancer. In this paper, we design two models based on the cool tip RF electrode and the multi-hook RF electrode.

The two models are based on a numerical finite element analysis to compute the distribution of heat and electric potential inside the damaged and surrounding tissue during an RF ablation. These two models simulate the ablation of large tumors while trying to reduce the damage to healthy cells in less time. This paper is organized as follows: First, an explanation of the methodology of the proposed model, including its equations. Second, the results and discussions of the proposed model show the differences in results between our model and others, and finally, the conclusions.

## II. METHODS

We used COMSOL 5.4 software [31] to implement FEM to develop two models (cool-tip and multi-hook RF electrode models). Each model is included within a cylindrical domain that contains both the liver domain and an electrode domain. The radius of the base of the cylinder is 5 cm, and the height is 12 cm.

### A. FEM Modeling

1) *Liver domain*: The liver tissue is the main tissue capable of generating glucose from lactate, glycerol, and amino acids (mainly alanine from muscle). It is located in the right upper quadrant of the abdominal cavity, rests just below the diaphragm, to the right of the stomach, and overlies the gallbladder. The liver domain includes everything surrounding the electrode: blood, blood vessels, and liver tissues.

#### 2) Electrode domains

a) *A cool-tip RF electrode domain*: It is distinguished by its excellent resistance to rust and corrosion as well as by its high flexibility, which avoids breakage or distortion, both of which pose risks to the patient's safety. The proposed model disregards the inner tube for brine injection responsible for cooling.

The cool tip RF electrode domain, consists of an electrode domain and a trocar domain as shown in Fig. 1 [32].

The RF electrode consists of:

- An insulated stainless-steel trocar with a diameter of 0.73 mm and a height of 12 cm.
- A nickel-titanium alloy (nitinol) electrode with a height of 3 cm and a diameter of 0.73 mm with a tip in a cone.

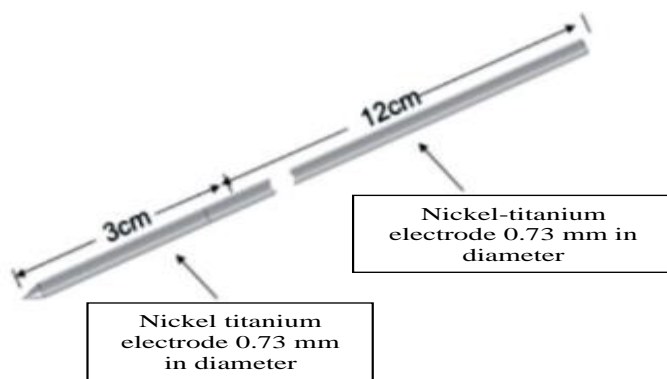


Fig. 1. A simplified model shows the design of the electrode [32].

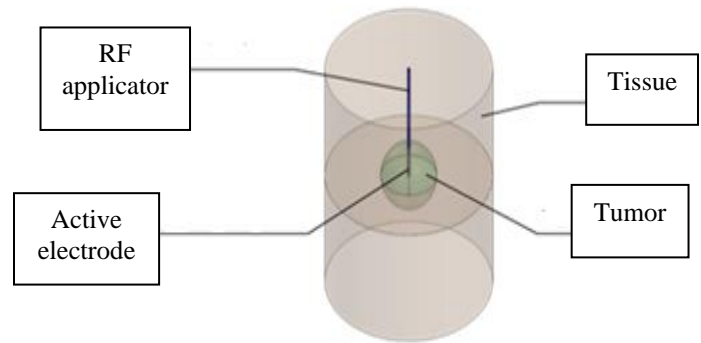


Fig. 2. A simple model that shows how a RF electrode works to ablate the tumor.

Fig. 2 explains the issue, where the tissue area is modelled as a cylindrical volume with the active electrode inserted into the tissue. The tissue area consists of the tumor in a spherical shape surrounded by healthy tissue inside the cylinder.

In our study, we set the power equal to 15 watts and assumed that the tumor was spherical, as 56% of liver cancer patients had spherical tumors [33]. We also assumed that the tumor size was 16 cm<sup>3</sup>, which is the largest volume that has been recorded for a liver tumor in Egypt in the last ten years [34].

#### a) A multi-hook RF electrode domain

Fig. 3 shows the design of multi-hook electrodes.

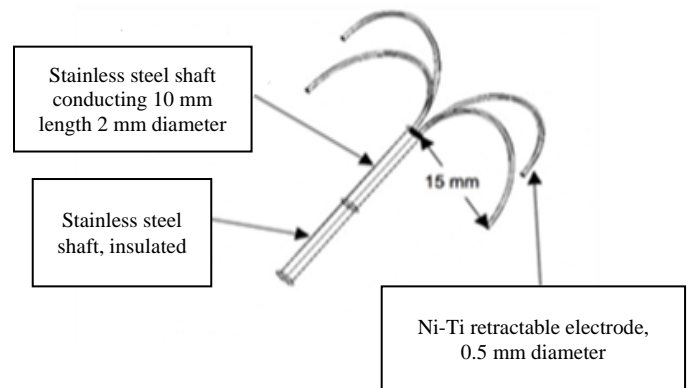


Fig. 3. Simplified model shows design of the electrode.

The multi-hook RF electrode consists of:

- Insulated stainless steel trocar with a height of 10 cm and a diameter of 2 mm.
- A stainless-steel electrode measures 1 cm in height and 2 mm in diameter, with a nickel-titanium alloy with a 1.5 cm major diameter and 0.5 mm minor diameter tip in a hook.

In our model, the initial conditions and boundary conditions of the simulation model are defined. We set the multi-hook RF electrode as the voltage source and set power equal to 15 watts, while we defined the liver boundary as the ground.

3) *Thermal and electrical properties:* Table I lists the different properties at 500 kHz used in our model to describe the gold (Au), the silver (Ag), the platinum (Pt), the nickel-titanium alloy (Nitinol), artery wall, and blood. The table shows the relative permittivity, density, and melting point values of the materials in the table are close. On the other hand, the table shows a clear variance in thermal and electrical conductivity values. So, we chose these two characteristics to study their effect on the ablation volume. During the ablation process, the tissue temperature increases. Changes in temperature lead to changes in the electrical conductivity of biological tissues [35]. The changes that occurred in the electrical conductivity of the tissues within the model are described as follows:

$$\delta(T) = \delta_0[1 + k_1\Delta T] \quad (1)$$

where:

$\delta_0$  is the initial electrical conductivity at the reference temperature (30°C),  $k_1$  is the temperature coefficient,  $\Delta T$  [°C] is the temperature difference from the initial reference temperature.

The linear electrical conductivity of the model is determined by.

$$\delta(T) = 0.155[1 + 0.0265\Delta T] \quad (2)$$

Expect the artery wall, and the equation is.

$$\delta(T) = 0.221[1 + 0.0265\Delta T] \quad (3)$$

Both equations are bounded at 30–80°C.

#### 4) Mathematical equations

a) *Electric field:* The electro-magnetic problem has been solved using a simplified version of Maxwell's equation that uses the quasistatic approximation because the displacement current is so minimal relative to the resistive current at the frequency range employed in the RFA technique (450–550 kHz). As a result, the generalized Laplace equation has been used to calculate the electric field.

$$\nabla \cdot (\sigma \Delta V) = 0 \quad (4)$$

Where:

$\sigma$  is (S/m) the electrical conductivity and V is the electric potential (V). The current study takes into account the tissue's electrical conductivity, which is temperature-dependent and increases linearly (10% per °C).

In the model tissues, the losses due to heat contact with the tissues are not significant, so we neglected the losses in our model [39]. The electric field E (V/m) is calculated from the equation.

$$E = -\nabla V \quad (5)$$

where  $\nabla$  is the operating factor. Then, current density J (A/m<sup>2</sup>) is calculated from

$$J = E/\rho e \quad (6)$$

where  $\rho e$  is the electrical resistivity of the material ( $\Omega$  m)

b) *Bioheat equation:* Calculations of the temperature distributions were made using FEM. By resolving the heat transfer equation (Equation 7), where we ignored the metabolic heat production  $Q_m$  and blood perfusion heat loss  $Q_b$  due to their comparatively small magnitudes, we were able to determine the temperature distribution throughout the tissue:

Bioheat transfer due to RF current can be mathematically explained by the following Pennes equation [40].

$$\rho c \partial T / \partial t = \nabla \cdot (k \nabla T) - \rho_b c_b \omega_b (T - T_b) + Q_m + JE + Q_b \quad (7)$$

$$JE = \sigma |\nabla V|^2 \quad (8)$$

$$Q_b = \omega_b c_b (T_b - T) \quad (9)$$

where  $\rho$  is the density of tissue (kg/m<sup>3</sup>),  $c$  is the specific heat capacity of the tissue (J/kg/K),  $k$  is the tissue thermal conductivity (W/m/K),  $\omega_b$  is the blood perfusion rate (1/s),  $Q_m$  is the volumetric heat produced by the metabolism (W/m<sup>3</sup>),  $JE$  is joule heating that represents heat generated from RF, where  $\sigma$  is the conductivity with a unit of  $S m^{-1}$ ,  $V$  is the voltage impressed on the electrode,  $Q_b$  is the volumetric heat produced by radiofrequency heating (Wm<sup>-3</sup>) calculated using Equation (2),  $T_b$  is the core blood temperature (supposed to be 37°C), because the blood vessels effect was discarded in this study then we can assume  $T_b$  and T are equal then  $Q_b$  was set to zero Wm<sup>-3</sup>

#### c) Boundary conditions

The boundary conditions are:

temp = normal body temperature (37 °C)

voltage = 0 V.

TABLE I. DIFFERENT PROPERTIES OF THE MODEL [36-38]

Material	Electrical conductivity $\sigma$ (S/m)	Thermal conductivity (W/ m. K)	Relative permittivity ( $\epsilon_r$ )	Density (kg/m3)	Melting point (°C)
Gold (Au)	44.2×10 <sup>6</sup>	315	1.143	1060	1064
Silver (Ag)	62.1×10 <sup>6</sup>	429	1.2	999	961.8
Platinum (Pt)	9.3×10 <sup>6</sup>	69.1	2.7	2190	1768
Nickel-titanium alloy (Nitinol)	2.4×10 <sup>6</sup>	18	86	6450	1650
Normal liver tissues	0.333	0.512	44.9	1060	-
Normal tumor tissues	0.1168	0.552	60.2	999	-

The complete FEM model's initial voltage and temperature have been taken into account to be 0 V and 37°C, respectively. The cylinder domain's bottom surface has been adjusted to 0 V to mimic a dispersive ground pad.

In the FEM model, an RF generator-controlled variable voltage source has been added to the electrode borders. All of the FEM model's remaining outside boundaries has been subjected to an electrical insulation boundary condition. For ten minutes, all of the numerical simulations of temperature-controlled RFA in various tissues were run. Furthermore, multiple temperature values were ranging from 50°C to 60°C.

d) *Ablated tissue*: The ablation area (i.e., where cell death occurs), it depends on the following factors:

- Damage time is the amount of time required to destroy the tumor-infected, damaged cell, equaling 10 minutes.
- The ablated temperature, equal to 50 °C, is the temperature required to destroy tumor-infected, damaged cells.
- Enthalpy change is the amount of heat evolved or absorbed in liver tissues (0 J/kg).

### III. RESULTS AND DISCUSSION

Cool-tip RF electrodes and multi-hook electrode ablation in our study depended on three materials: gold, silver, and platinum. This study analyzed these materials to provide guidance in clinical practice.

#### A. Cool-tip RF Electrode

Table II shows the ablation volumes due to the used electrode materials at different ablation durations.

TABLE II. RESULTS FROM SIMULATION WITH A COOL-TIP RF ELECTRODE

Time (min)	Electrode material	Ablation volume(cm <sup>3</sup> )
1	Nitinol	2.5
	Pt	4.7
	Ag	6.4
	Au	7.6
2.5	Nitinol	3.7
	Pt	6.8
	Ag	8.4
	Au	9.5
5	Nitinol	5.1
	Pt	8.4
	Ag	10.1
	Au	11.3
7.5	Nitinol	6.2
	Pt	10.3
	Ag	12
	Au	13.3
10	Nitinol	7.4
	Pt	12.2
	Ag	13.9
	Au	15.2

The ablation volume of tumor results output from gold, silver, platinum, nickel-titanium alloys, and electrodes respectively is shown in Fig. 4, 5, 6, and 7, respectively.

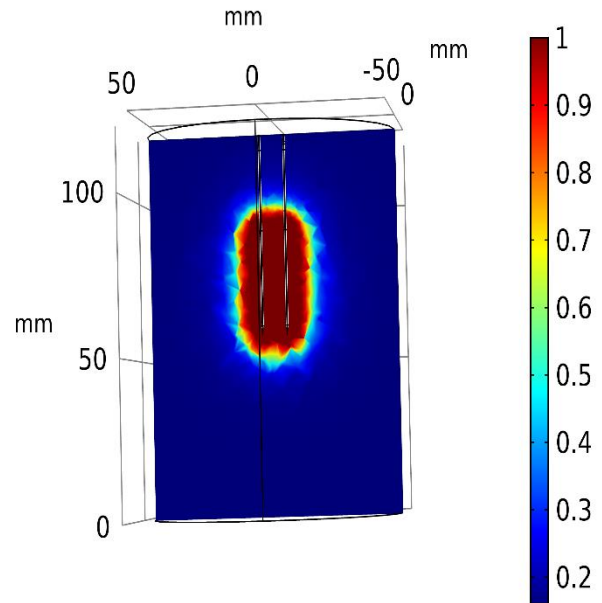


Fig. 4. View of the fraction of damage and the ablation volume resulting from using a cool-tip RF electrode made of Gold after ten minutes.

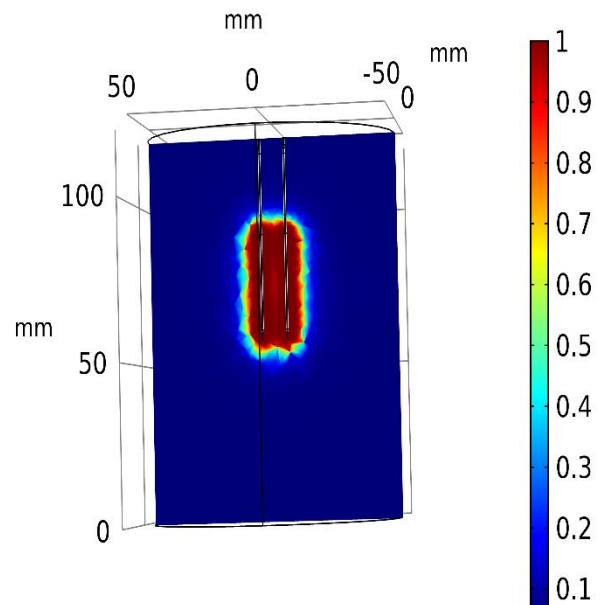


Fig. 5. View of the fraction of damage and the ablation volume resulting from using a cool-tip RF electrode made of Silver after ten minutes.

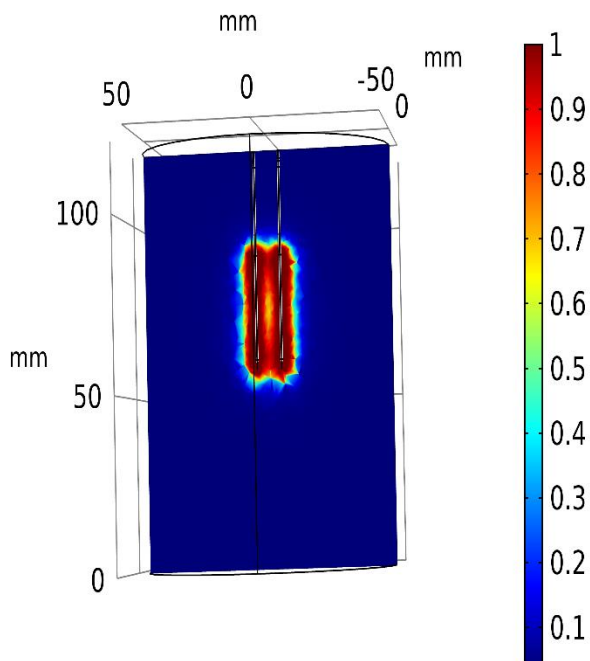


Fig. 6. View of the fraction of damage and the ablation volume resulting from using a cool-tip RF electrode made of Platinum after ten minutes.

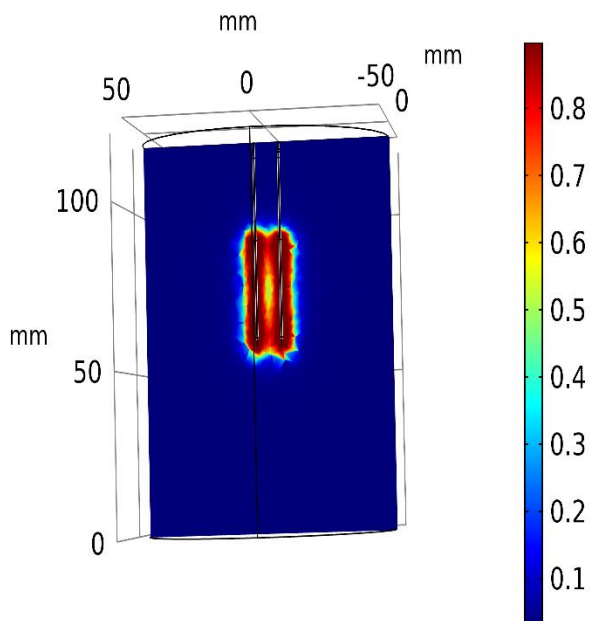


Fig. 7. View of the fraction of damage and the ablation volume resulting from using a cool-tip RF electrode made of Nickel-Titanium alloy after ten minutes.

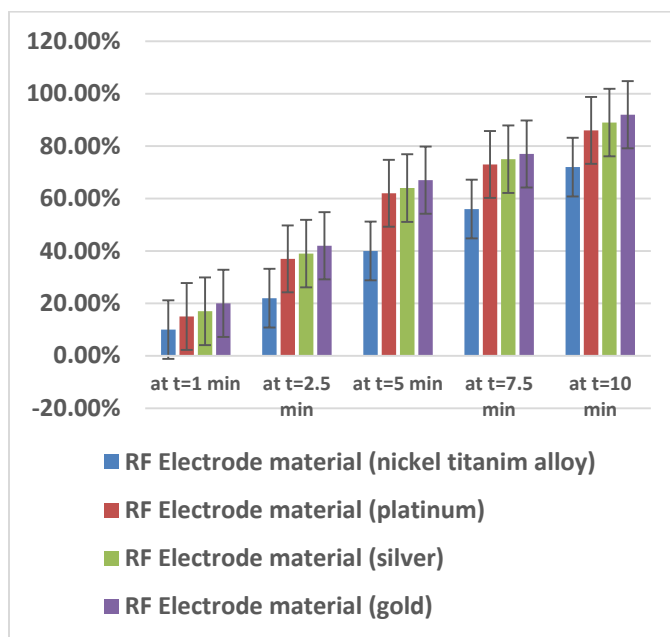


Fig. 8. Comparison of percentage of ablation volume for cool-tip RF electrodes made of Gold, Silver, Platinum, and Nickel-Titanium alloy.

The simulation results and the percentage of ablation achieved from gold, silver, platinum, and nickel-titanium alloy are shown graphically in Fig. 8.

### B. Multi-hook RF Electrode

Table III shows the results for the simulations with the new design of multi-hook RF electrode shown in Table III. At the same time and with the same power, there was an appreciable difference in the ablation volume between the different electrode materials.

TABLE III. SIMULATION RESULTS WITH A MULTI-HOOK RF ELECTRODE

Time (min)	Electrode material	Ablation volume(cm <sup>3</sup> )
1	Nitinol	2.6
	Pt	4.8
	Ag	6.5
	Au	7.8
2.5	Nitinol	4
	Pt	6.9
	Ag	8.5
	Au	9.8
5	Nitinol	5.3
	Pt	8.5
	Ag	10.2
	Au	11.7
7.5	Nitinol	6.5
	Pt	10.6
	Ag	12.4
	Au	13.5
10	Nitinol	7.9
	Pt	12.3
	Ag	14.3
	Au	15.4

The ablation volume of tumor results output from gold, silver, platinum, and nickel-titanium alloy RF electrodes, respectively, is shown in Fig. 9, 10, 11, and 12, respectively.

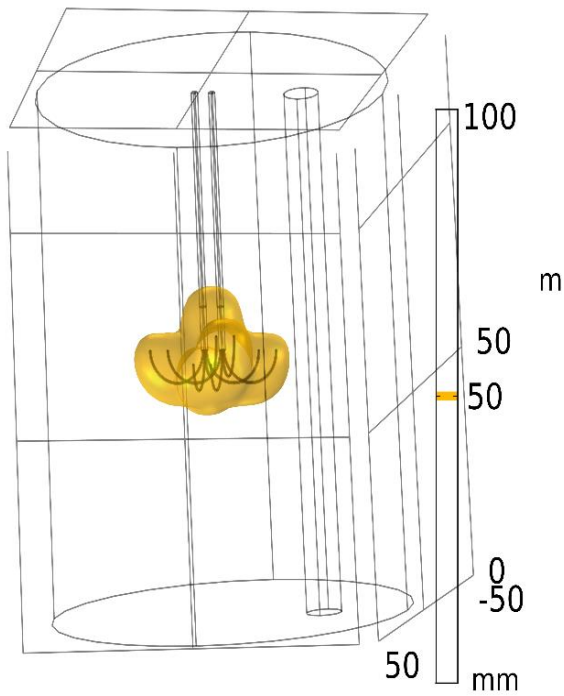


Fig. 9. View of temperature distribution and the ablation volume resulting from using a multi-hook RF electrode made of Gold after ten minutes.

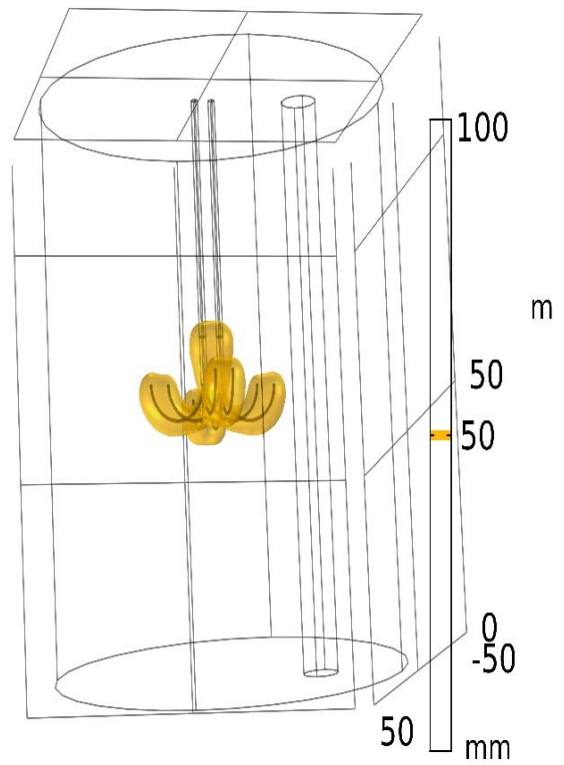


Fig. 11. View of temperature distribution and the ablation volume resulting from using a multi-hook RF electrode made of Platinum after ten minutes.

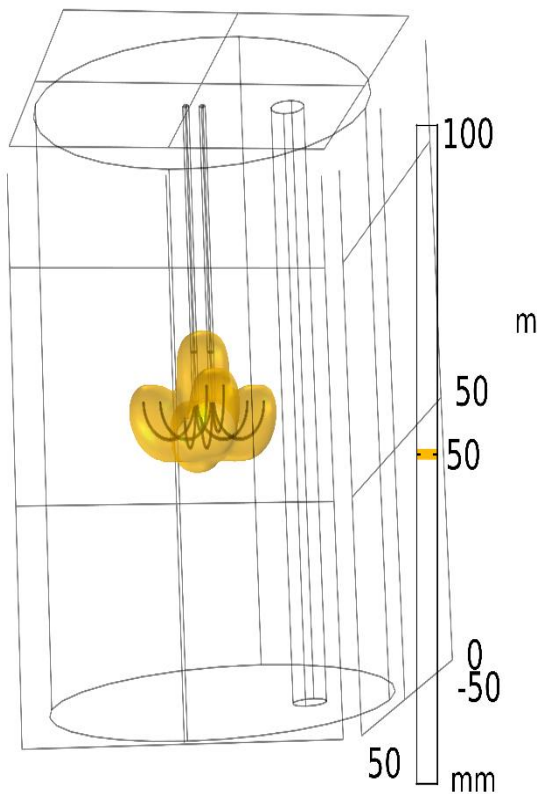


Fig. 10. View of temperature distribution and the ablation volume resulting from using a multi-hook RF electrode made of Silver after ten minutes.

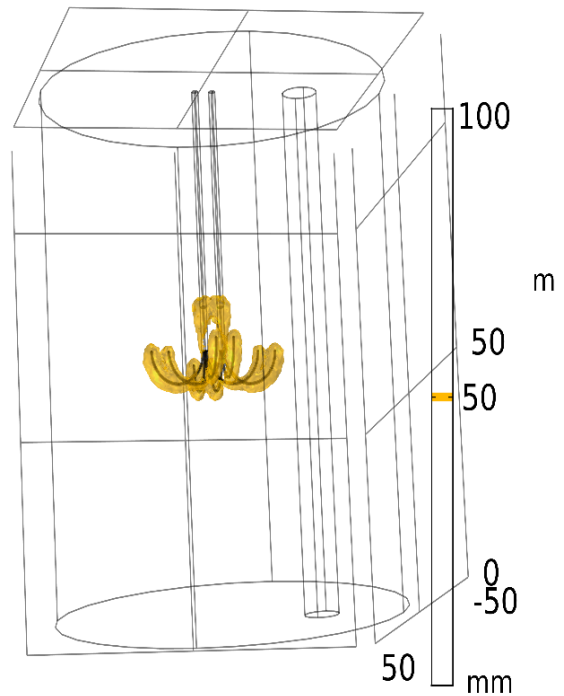


Fig. 12. View of temperature distribution and the ablation volume resulting from using a multi-hook electrode made of Nickel-Titanium alloy after ten minutes.

The simulation results and the percentage of ablation achieved from gold, silver, platinum, and nickel-titanium alloy are shown graphically in Fig. 13.

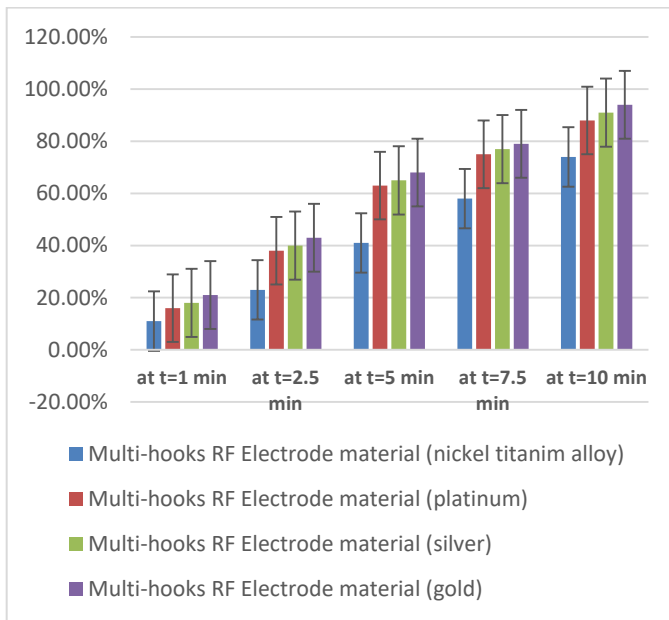
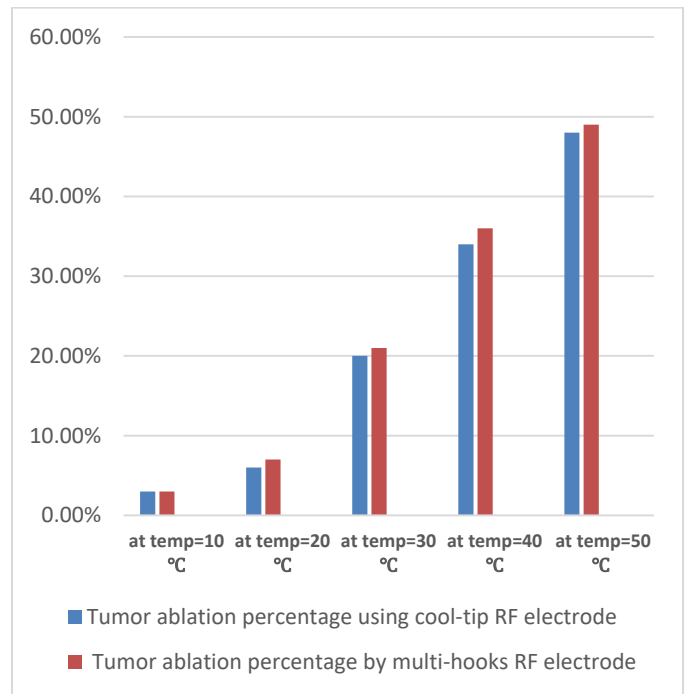


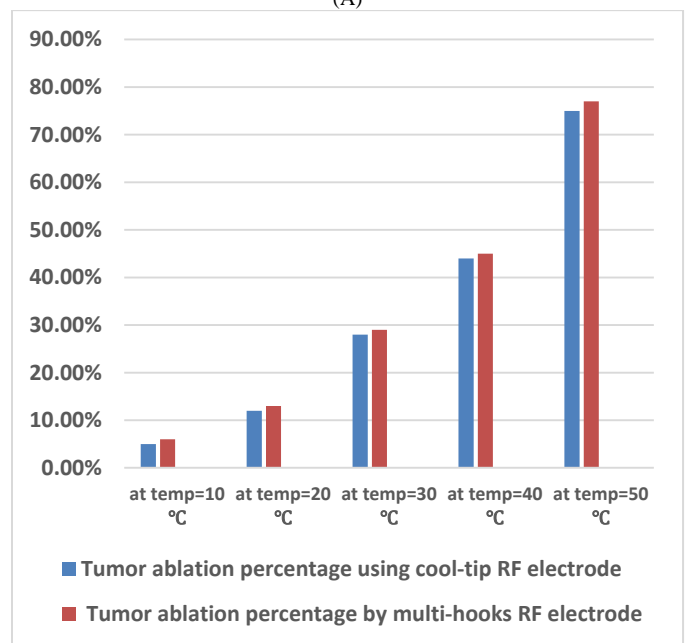
Fig. 13. Comparison of percentage of ablation volume for multi-hook RF electrodes made of Gold, Silver, Platinum, and Nickel-Titanium alloy.

In this paper, we designed a computer model to investigate whether gold, silver, and platinum RF electrodes create larger volume ablation than nickel-titanium alloy electrodes using the same conditions. As shown from the results, for RF and multi-hook electrodes, there is a difference in the percentage of ablation using RF electrodes made of gold, silver, and platinum materials compared to using an electrode made of a nickel-titanium alloy by an amount ranging from five percent to ten percent after a minute, while the difference in the percentage of ablation ranged from 14% to 30%. On the other hand, the ablation rate with the multi-hook electrode is 2% greater than the ablation rate with the cool-tip RF electrode in all kinds of materials used as the multi-hook electrode covers a larger area due to the design difference between the two electrodes.

Previous theoretical research looked into the impacts of using Pt, Ag, and Au as the electrode material for RF ablation rather than nitinol since these metals have higher thermal and electrical conductivities. According to the results, a different amount of ablation was produced, with each of the four electrode materials. Using varied powers at various intervals, we investigated in this study whether Pt, Ag, and Au electrodes produced thermal ablation more successfully than nitinol electrodes. In all simulations, we proved that as the temperature rose, more power was given via the Pt, Ag, and Au electrodes, leading to greater maximum tissue temperatures and a larger ablation volume, as depicted in Fig. 14, 15, and 16. We discovered that the four electrodes' ablation volumes varied.



(A)



(B)

Fig. 14. Comparison of the percentage of ablation volume for cool-tip and multi-hook RF electrodes, (A) made of Nitinol, and (B) made of Platinum.

Therefore, differences in the thermal and electrical conductivities of RF electrode material offer solutions in certain clinical situations. Our modelling succeeded in reducing time and reducing patient pain.



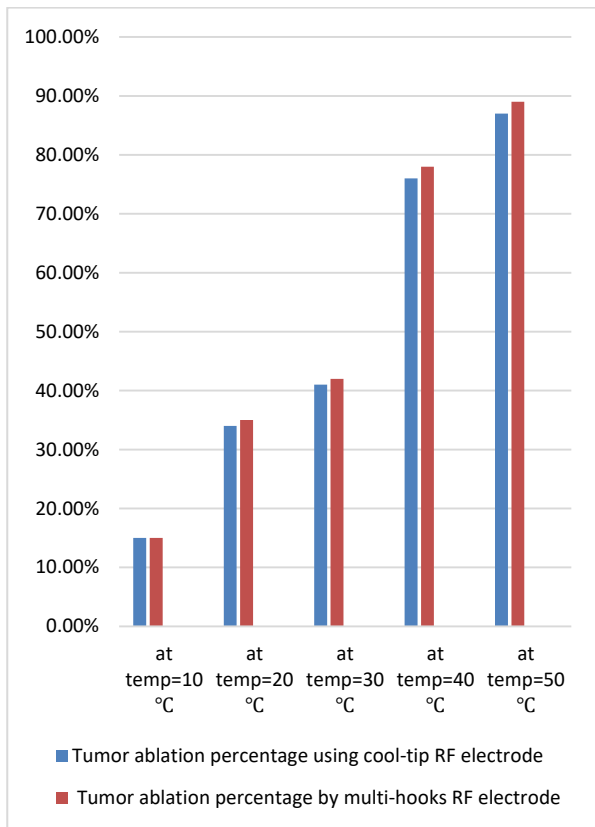


Fig. 15. Comparison of the percentage of ablation volume for cool-tip and multi-hook RF electrodes made of Silver.

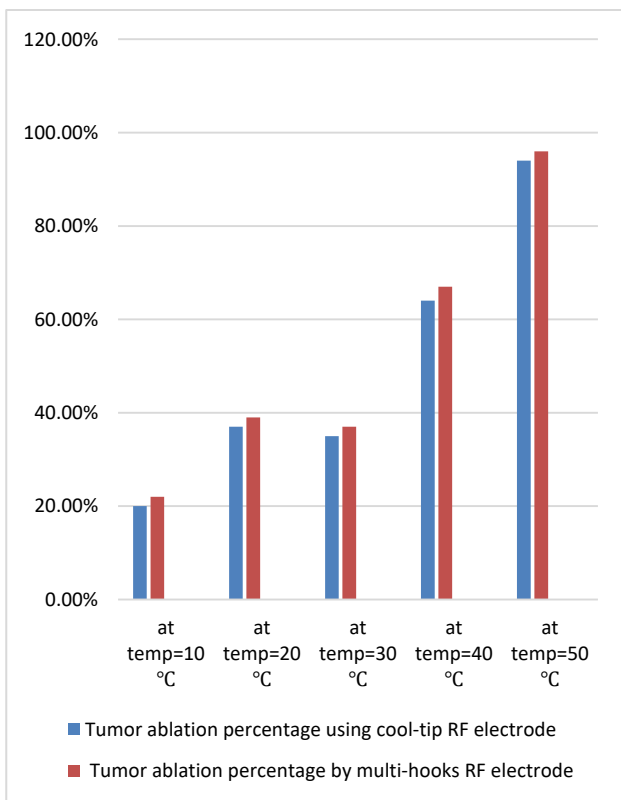


Fig. 16. Comparison of the percentage of ablation volume for cool-tip and multi-hook RF electrodes made of Gold.

#### IV. CONCLUSION

The results show that using gamma titanium instead of nickel-titanium achieved an increase in the ablation volume because it has higher electrical and thermal conductivity than nickel-titanium alloy. Also, the results show the effect of tuning parameters (ablation power, ablation time and design of the electrode) on increasing the ablation volume and decreasing the ablation time. The results show that the electrode design is the most important tuning parameter because with the change of design from the reference model to our developed model, the effect and success of the rest of the tuning parameters appeared. That helps to reduce the patient's pain and increase the accuracy. Future work can study the effect of the RF multi-hooks electrodes on large tumors by using the same tuning parameters and compare it with a RF cool-tip electrode.

#### REFERENCES

- [1] Ferlay, J., et al., Cancer statistics for the year 2020: An overview. *International Journal of Cancer*, 2021. 149(4): p. 778-789.
- [2] Rashed, W.M., et al., Hepatocellular Carcinoma (HCC) in Egypt: A comprehensive overview. *Journal of the Egyptian National Cancer Institute*, 2020. 32(1): p. 1-11.
- [3] Cao, W., et al., Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. *Chinese Medical Journal*, 2021. 134(07): p. 783-791.
- [4] Sun, Q., et al., Survival analysis following microwave ablation or surgical resection in patients with hepatocellular carcinoma conforming to the Milan criteria. *Oncology Letters*, 2020. 19(6): p. 4066-4076.
- [5] Jarosova, J., et al., Endoscopic radiofrequency ablation for malignant biliary obstruction. *World Journal of Gastrointestinal Oncology*, 2021. 13(10): p. 1383.
- [6] Khattab, A., Regional distribution of lung ventilation during high-frequency jet ventilation in anesthetized patients. 2022, Wien.
- [7] Ghosh, R., Computational Modelling and Simulation of RF Induced Heating of Biological Tissue. 2020, State University of New York at Buffalo.
- [8] Chen, Q., Model-Based pre-operational plan optimization of hepatic tumor radiofrequency ablation. 2015: Northeastern University.
- [9] Lee, M.W., et al., Updated 10-year outcomes of percutaneous radiofrequency ablation as first-line therapy for single hepatocellular carcinoma < 3 cm: emphasis on association of local tumor progression and overall survival. *European radiology*, 2020. 30(4): p. 2391-2400.
- [10] Kim, N., et al., Retrospective analysis of stereotactic body radiation therapy efficacy over radiofrequency ablation for hepatocellular carcinoma. *Radiotherapy and Oncology*, 2019. 131: p. 81-87.
- [11] Fang, Z., et al., Design of a novel electrode of radiofrequency ablation for large tumors: A finite element study. *Journal of Engineering and Science in Medical Diagnostics and Therapy*, 2018. 1(1).
- [12] Rossmann, C. and D. Haemmerich, Review of temperature dependence of thermal properties, dielectric properties, and perfusion of biological tissues at hyperthermic and ablation temperatures. *Critical Reviews™ in Biomedical Engineering*, 2014. 42(6).
- [13] Chang, I., Finite element analysis of hepatic radiofrequency ablation probes using temperature-dependent electrical conductivity. *Biomedical engineering online*, 2003. 2(1): p. 1-18.
- [14] Eliaz, N., Corrosion of metallic biomaterials: A review. *Materials*, 2019. 12(3): p. 407.
- [15] Jagota, V., A.P.S. Sethi, and K. Kumar, Finite element method: an overview. *Walailak Journal of Science and Technology (WJST)*, 2013. 10(1): p. 1-8.
- [16] Toyohara, R., et al., Finite element analysis of load transition on sacroiliac joint during bipedal walking. *Scientific reports*, 2020. 10(1): p. 1-10.

- [17] Alemayehu, D.-B. and Y.-R. Jeng, Three-dimensional finite element investigation into effects of implant thread design and loading rate on stress distribution in dental implants and anisotropic bone. *Materials*, 2021. 14(22): p. 6974.
- [18] Segura Félix, K., et al., Computational FEM Model and Phantom Validation of Microwave Ablation for Segmental Microcalcifications in Breasts Using a Coaxial Double-Slot Antenna. *BioMed Research International*, 2021. 2021.
- [19] Mulier, S., et al., Radiofrequency ablation with four electrodes as a building block for matrix radiofrequency ablation: Ex vivo liver experiments and finite element method modelling. Influence of electric and activation mode on coagulation size and geometry. *Surgical Oncology*, 2020. 33: p. 145-157.
- [20] Singh, S. and R. Repaka, Numerical study to establish relationship between coagulation volume and target tip temperature during temperature-controlled radiofrequency ablation. *Electromagnetic biology and medicine*, 2018. 37(1): p. 13-22.
- [21] Mertyna, P., et al., Radiofrequency ablation: the effect of distance and baseline temperature on thermal dose required for coagulation. *International journal of hyperthermia*, 2008. 24(7): p. 550-559.
- [22] Zorbas, G. and T. Samaras, Simulation of radiofrequency ablation in real human anatomy. *International Journal of Hyperthermia*, 2014. 30(8): p. 570-578.
- [23] Singh, S. and R. Repaka, Quantification of thermal injury to the healthy tissue due to imperfect electrode placements during radiofrequency ablation of breast tumor. *Journal of Engineering and Science in Medical Diagnostics and Therapy*, 2018. 1(1).
- [24] Diebold, A., et al., Electrowetting-actuated liquid metal for RF applications. *Journal of Micromechanics and Microengineering*, 2017. 27(2): p. 025010.
- [25] Zhao, P., et al., RF Performance Benchmarking of TSV Integrated Surface Electrode Ion Trap for Quantum Computing. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 2021. 11(11): p. 1856-1863.
- [26] Safavi, A. and M. Tohidi, Microwave-assisted synthesis of gold, silver, platinum and palladium nanostructures and their use in electrocatalytic applications. *Journal of Nanoscience and Nanotechnology*, 2014. 14(9): p. 7189-7198.
- [27] Pruneanu, S., et al., Electro-catalytic properties of graphene composites containing gold or silver nanoparticles. *Electrochimica Acta*, 2013. 89: p. 246-252.
- [28] Yamada, M., M. Foote, and T.W. Prow, Therapeutic gold, silver, and platinum nanoparticles. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, 2015. 7(3): p. 428-445.
- [29] Im, C. and J.-M. Seo, A review of electrodes for the electrical brain signal recording. *Biomedical Engineering Letters*, 2016. 6(3): p. 104-112.
- [30] Kumph, M., et al., Operation of a planar-electrode ion-trap array with adjustable RF electrodes. *New Journal of Physics*, 2016. 18(2): p. 023047.
- [31] Multiphysics, C. COMSOL Multiphysics® 5.4. COMSOL Multiphysics. 2021; Available from: <https://www.comsol.com/>.
- [32] Xu, L., et al., Simulation of multi-probe radiofrequency ablation guided by optical surgery navigation system under different active modes. *Computer Assisted Surgery*, 2016. 21(1): p. 107-116.
- [33] Mahnič-Kalamiza, S. and D. Miklavčič, Scratching the electrode surface: Insights into a high-voltage pulsed-field application from in vitro & in silico studies in indifferent fluid. *Electrochimica Acta*, 2020. 363: p. 137187.
- [34] Attia, M.S., et al., A new method for early diagnosis of liver cancer using a biosensor embedded in an alginate polymer thin film. *Journal of Materials Chemistry C*, 2022. 10(16): p. 6464-6472.
- [35] Ng, E. and M. Jamil, Parametric sensitivity analysis of radiofrequency ablation with efficient experimental design. *International journal of thermal sciences*, 2014. 80: p. 41-47.
- [36] Liang, C., et al., Modeling and analysis of thermal characteristics of magnetic coupler for wireless electric vehicle charging system. *IEEE Access*, 2020. 8: p. 173177-173185.
- [37] Zhang, C., et al., Coupled mechanical-electrical-thermal modeling for short-circuit prediction in a lithium-ion cell under mechanical abuse. *Journal of Power Sources*, 2015. 290: p. 102-113.
- [38] Nonneman, J., et al. Quality Assessment of a 2D FE Based Lumped Parameter Electric Motor Thermal Model Using 3D FE Models. in 2020 International Conference on Electrical Machines (ICEM). 2020. IEEE.
- [39] Mi, Y., et al., Multi-parametric study of temperature and thermal damage of tumor exposed to high-frequency nanosecond-pulsed electric fields based on finite element simulation. *Medical & biological engineering & computing*, 2017. 55(7): p. 1109-1122.
- [40] Ghazanfarian, J., R. Saghatchi, and D. Patil, Implementation of smoothed-particle hydrodynamics for non-linear Pennes' bioheat transfer equation. *Applied Mathematics and Computation*, 2015. 259: p. 21-31.

# A Light-weight Authentication Scheme in the Internet of Things using the Enhanced Bloom Filter

Xiaoyan Huo

Information Construction and Management Center, Jiaozuo University, Jiaozuo, Henan, 454003, China

**Abstract**—Authenticated key exchange mechanisms are critical for security-sensitive Internet of Things (IoT) and Wireless Sensor Networks (WSNs). In this area, the Bloom Filter (BF) plays a crucial role directly and indirectly, which has a significant advantage in space and time. Light-weight input authentication is one of the most challenging tasks in IoT. Weak or inefficient defense algorithms can allow fake information to enter the system, share information, send unnecessary messages, and reduce network efficiency. The utilization of an augmented Bloom filter for creating an authentication prominent called En-route Authentication Bitmap (EAB) has a substantial advantage over traditional methods that involve direct usage of Message Authentication Codes (MAC). This effective method of EAB picks the fake information almost accurately, thereby reducing the feeding attacks within not more than two steps taken by the attacker. EAB necessarily needs only a few bytes of bandwidth for efficient defense against at least ten forward steps of the adversary. Without hesitation, the Augmented Bloom filter and its components are becoming more common in network defense mechanisms.

**Keywords**—En-route authentication bitmap; message authentication codes; internet of things; bloom filter

## I. INTRODUCTION

As the Internet of Things (IoT) expands, it is poised to transform human lifestyles and release several monetary benefits [1, 2]. Security and trust issues present significant adoption limitations for the IoT [3]. Blockchain, a distributed and tamper-resistant ledger, maintains regular statistics data at only places and can deal with the information protection situation in IoT networks [4]. The IoT is an entirely novel paradigm combining technologies and elements that come from different strategies [5]. The real and virtual worlds constantly interact through embedded devices, communication technology, sensing technology, Internet protocol, pervasive computing, and ubiquitous computing [6]. The IoT vision relies on smart objects. Putting intelligence into everyday objects enables them to collect environmental data, interact with the physical world, and interconnect via the Internet with alternate facts and records [7]. New business opportunities will arise as devices become increasingly connected and records become increasingly plentiful. This enables delivering tangible advantages to individual residents, the economy, the environment, and society [8].

Designers are primarily concerned with the trade-off between cost, computation, and security [9]. Although end-to-end authentication can be validated by sharing keys between the source and target, accuracy is complicated halfway, where intermediate nodes cannot determine authenticity without

additional information [10]. In wireless broadcasting, promotional messages should be eliminated as much as possible [11]. Data passes through a node route every time. The relative nodes are instructed to remain still and observe the information even though those nodes are neither the target nor the transporters. Making things worse, the transfer packet data does not make use only the way where the neighbor nodes are present but also the neighbor region [12].

There is competition between the two-hop nodes for communication channels, similar to the hidden ones problem [13]. Hop-by-hop routing easily identifies entryways for the attackers to feed and take the system information [14]. The endpoint cannot be made to believe the take information by the adversary unless the two-end key is weak or unsafe [15]. Providing false messages in the pathway, redundant traffic creation, waste usage of computation power of the route, and retarding the efficiency of the network can usually be made simple by the attacker [16]. Using a network-wide authentication key is a natural way to solve this problem. However, the adjustment of one node can lead to the adjustment of the whole arrangement. It can be seen as the most effective way to follow  $\mu$ Tesla, in which the authentication key is updated frequently. Nevertheless, this method has its advantage in that network synchronization with standard timing error should be present. Synchronizing the timing is complicated if the route is extensive. Moreover, these approaches are suitable for central nodes, like base stations [17].

Attaching all Message Authentication Codes (MACs) with the source node to the payload for every router is another effective solution [18]. As the number of route nodes increases, the volume of the message gradually increases. These problems necessitate the need for lightweight en-route authentication schemes in multi-hop networks. This paper discusses a statistical way to solve the hop-by-hop message authentication problem between two ends. We efficiently examine data filtration in a routing way for bandwidth mitigation. En-route Authentication Route Bitmap (EAB) is a structure of the bloom filter method that is appended to appropriate messages as extra data. We propose a method for protecting the system from attack by injecting bogus data since the damage is contained in its immediate area. Slight numerical adjustments can improve efficiency and protection.

The challenges of IoT security are securing restrained devices, empowering devices, validating devices, updating devices, ensuring the confidentiality and reliability of data, managing web, mobile, and cloud functions, guaranteeing high availability, detecting vulnerabilities and incidents, handling

vulnerabilities, and forecasting security issues. To overcome the constraints in the IoT, we propose a system that can defend the Internet of Things from attack and implement a lightweight authentication. Bloom filters are primarily responsible for authentication. This paper implements an Augmented Bloom Filter. The EAB is inaccurate for mapping multiple MACs. EAB classifies false data during the first constrained hop instead of being dropped at the first sign. Therefore, the false negative rate adjacent to the adversary is satisfactory, but it quickly rises with increasing bound counts. This swapping drastically reduces authentication communication overhead.

Multihop wireless networks face challenges with lightweight en-route authentication. False data can be injected into a system in the absence of an efficient defense mechanism, causing redundant message forwarding and consuming node power. We create an authentication manifest using EAB, which differs from conventional approaches that directly use MACs. By filtering out false data with a high success rate, EAB makes injection attacks less likely to spread beyond one or two hops. While EAB spends a few bytes of bandwidth, it can effectively protect tens of hops along the forwarding path.

The rest of the paper is organized in the following manner. Preliminaries are presented in Section II. Basis design of the mechanism is discussed in Section III. A discussion of random padding improvements is presented in Section IV. Section V concludes the paper.

## II. RELATED WORK

Nianmin, Haifeng [19] proposed a method for ensuring data integrity. This one-level hash structure has three advantages over traditional hash trees: a lower computation overhead, a lower space overhead, and the ability to adjust the security level. The proposed method was compared with an efficient integrity checking scheme for its overhead and security. According to the evaluation results, the method has a lower overhead for most benchmarks.

Saravanan and Senthilkumar [20] present an enhanced bloom filter technique to represent large sets of data in a secure manner and apply it to distributed applications such as web caching and peer networks. Also, to limit the possibility of unauthorized access to the dataset by unauthorized intruders, the enhanced bloom filter is applied with an upper bound on the false-positive probability by increasing its capacity as the packet data size increases. As compared to the existing dynamic bloom filter approach, the experimental results show a 42.5% improvement in packet data security.

A scheme based on an improved counting bloom filter was developed by Wang, Wang [21] for securing authentication keys in heterogeneous sensor networks. By applying the Set theory to heterogeneous sensor networks, the counting bloom filter algorithm is improved to address the authentication key agreement problem. As demonstrated by experimental results, the proposed scheme has greater network scalability, lower communication costs, and can resist brute force attacks during a node capture.

Malhi and Batra [22] proposed an authentication framework based on pseudonyms that preserves privacy. An ID-based signature scheme is used for vehicle-to-RSU

communication, and a new digital signature scheme is designed for vehicular communications. The identity of the vehicle is revealed by multiple authorities in the event of revocation. Bloom filters are used to improve the signature verification scheme, which was implemented on a simulated environment to evaluate the results.

Mbarek, Sahli [23] developed an efficient authentication method to authenticate sensors utilizing the Bloom Filter. DoS attacks can be mitigated by using multiple MACs along with a Bloom filter in delayed key disclosure schemes. A set of scenarios was used to assess the feasibility of the protocol. Authentication protocol reduces false positives in Bloom Filters, according to the results.

By derived from the cryptographic permutation Xoodoo, Sateesan, Vliegen [24] proposed a new noncryptographic hash function, called Xoodoo-NC, for developing ultra-high-speed Bloom filters on FPGAs. The Xoodoo-NC hash function inherits the desired avalanche properties of Xoodoo and the low logical depth, resulting in an ultra-low-latency non-cryptographic hash function.

## III. PRELIMINARIES

### A. System Model

There are several computing nodes in IoT. These nodes can pick the links between themselves and their adjacent nodes and set up a connection with nodes that are far away by implementing a path through the nodes in the way. There may be several server nodes in existence containing effective resources that may include cloud computing and fog computing. Node A and Node B can be considered fellow nodes if the authentication key is shared. These authentications between fellow nodes that are end-to-end are certified with contributed keys through some systems. To share information between two end nodes, both nodes should possess a common key, or the sender should appoint the information to be delivered to both end nodes. The division of the path is considered a series of transmission systems. Using step-by-step verification as a benefit of end-to-end information sharing is mainly discussed here. Considering the routing paths to be constant, we assume it to be sensed by the node from the origin.

### B. Threat Model

End-to-end authentication usually assures the security and purity of the information. This security can be breached by exposing the pairwise key among the source and target nodes. Imposing false traffic and message authentication on the nodes' path and diminishing the network's efficiency is the attacker's main aim in this case. Here the adversary creates upstream packets or forge packets and spreads them in the route path. These can falsely act as the origin or pathway of some other fake data. Rejection of messages that come across the path is an alternative attack technique of the adversary. This type of attack naturally does not have the right solution until there is an effective connection between the relative nodes.

### C. Bloom Filters

Bloom Filters are probabilistic data structures with some errors. It is a simple hashing algorithm that requires very little

memory. Various research areas have been applied by Bloom Filter to boost performance. As a membership filter, it returns either "true" or "false". The term "true" can, however, refer to either a false positive or true positive. In the same way, a "false" can also be a true negative or a false negative. False positives and false negatives are errors of the Bloom Filter. However, the error is negligible and tolerable. Nevertheless, Bloom Filter is not suitable for many systems, such as real-time systems.

A very less false positive and null false negative is effectively obtained through the Bloom filter, which is greatly useful in representing a pack of inputs to fulfill the above criteria. Firstly, every bit used in the Bloom filter remains unmarked, and  $k$  is set to be stochastic hash functions.  $H = \{H^1, \dots, H^k\}$  where the range is from  $\{0, \dots, m-1\}$ . Representation of the set  $S = \{s_1, \dots, s_n\}$  of a number of inputs ( $n$ ),  $n$  and  $H_i(s_j)$  are set to 1 where  $1 \leq j \leq n$  and  $1 \leq i \leq k$ . Where every input  $s_i$  are set to 1 to the bitmap, the chances for a particular bit to be marked is  $p = 1 - (1 - 1/m)^{kn}$ . It results in nearly  $mp = (1 - (1 - 1/m)^{kn})$  bits in the map being marked averagely. The discussed method is effectively used to create a Bloom filter, where  $Bitmap^{bloom} \leftarrow H, S$ .

In order to evaluate a new input  $x$  in  $S$ , a mark for every  $hi(x)$  bit is considered. If everything is marked, then it can be definite that  $x$  is input in  $S$ . Here, the only flow along with the small false positive chance  $f = (1-p)^k$  is probable, based on the value of  $m/n$  and  $k$ . If anything is left unmarked,  $x$  is definitely not present in  $S$ .

Border and Mitzenmacher coined the Bloom filter. Using a list or sets at premium spaces, using the Bloom filter is relevant, where it is possible to lower the severity of false positives. Augmented Bloom filters are more efficient when compared to other set representation data structures such as hash tables, binary search trees, linked lists, or simple arrays. Here we use a bloom filter array for a defense mechanism. The augmented bloom filter proceeded with two main steps, insertion operation and membership testing. Firstly, the elements in the set  $S$  should be inserted into the bloom filter array of  $m$  bits. Let the elements be  $x_1, x_2, x_3, \dots, x_n$ , where  $n$  is the number of elements. This makes it easier for members to test whether the element is present in the set or not.

#### D. Insertion Operation

The elements in the set  $S$  are added to the bloom filter array after yielding corresponding hash keys of individual elements using hash functions such as MD5. Here we are using a hash function named MD5 algorithm. The usage of the hash function greatly reduces the occurrence of false positives. In contrast, using a single hash function to insert elements in the bloom filter array leads to a high possibility of false positives. There will be no false negative outputs either by multiple hash functions or by implementing the hash function differently into the bloom filter.

Usually, the bloom filter implements three hash function algorithms to obtain three different key values. Here we will utilize a single hash function, but a slightly different method is preferred. Initially, divide the input value into three parts and then hashed by using the MD5 algorithm individually, which

can be represented as  $hk_1(x_1), hk_2(x_2), hk_3(x_3)$ , where  $k_1, k_2,$  and  $k_3$  are the hash functions. In order to generate hexadecimal values using the MD5 algorithm, the following steps must be sequentially applied.

1) *Appending padding bits:* In the first step, the element  $x$ , initially a URL, is converted to bits. By adding a single '1' bit at the end of the 'b' bit message, the message becomes divisible by 448 or 512.

2) *Appending length:* By adding a 64-bit representation, an output multiple of 448 can be converted into a multiple of 512.

3) *Buffer initialization:* This step divides the b-bit result from the preceding step into four 32-bit registers (A, B, C, D). A 128-bit message digest is derived using these registers.

4) *Processing the message:* Four auxiliary functions are applied to the message, and various processing steps produce the required output. The plain text is then converted into cipher text, resulting in the message digest.

We convert the obtained hexadecimal values into binary values, which can finally be inserted into the bloom filter, where the bits are initially zero. After adding the elements in the set, those '0' bits on the corresponding positions will eventually become '1' based on the inserted element.

#### E. Membership Testing

We check their similarity, ensuring it is already present in the array. Hence the bits need not be stored in the Augmented Bloom filter array. If not, we should add those bits to the Augmented Bloom filter array after checking whether the current bit is 1. If the relevant Augmented Bloom filter array is 0, it changes to 1. Otherwise, nothing is changed in membership testing.

1) *Augmented bloom filter array:* The Bloom filter is a space-efficient probabilistic statistics structure for determining whether elements belong to sets or not. They are astonishingly honest: take an array of  $k, n, m,$  and  $p$ . Both are examined for the use of hash features. By setting all bits, the element possibly now exists via a false positive rate of  $p$ ; if the number of the bits is not specified, the element genuinely does not exist. Bloom filters locate an in-depth type of usage, counting and tracking articles that one has to examine, accelerating Bitcoin clients, detecting malicious links, and improving cache performance. This will help us select the most reliable size for our filter. Fig. 1 to 3 show graphs among  $p$  versus  $n, m,$  and  $k$ .  $n$  refers to the number of items within the filter,  $p$  stands for the probability of false positives, the fraction between zero and one or various representing 1-in- $p$ ,  $m$  signifies the number of bits in the filter out, and  $k$  shows the number of hash functions. These values are computed by Eq. 1-4.

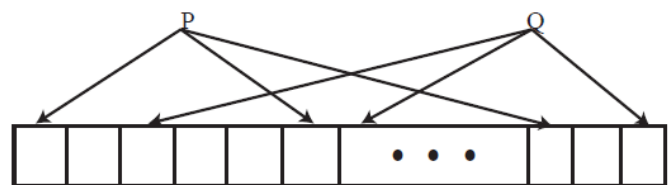


Fig. 1. Dynamic bloom filter.

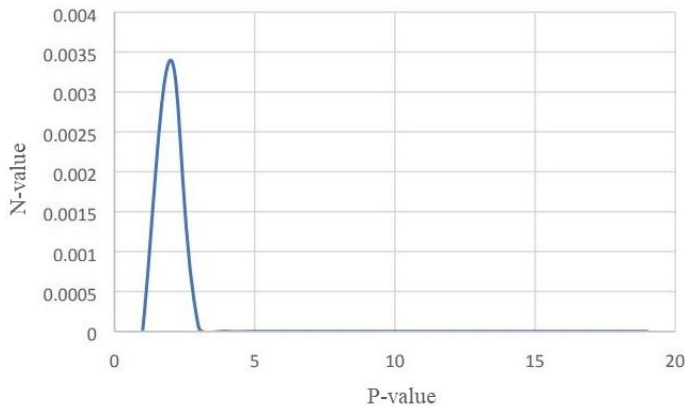


Fig. 2. Graph p vs. n.

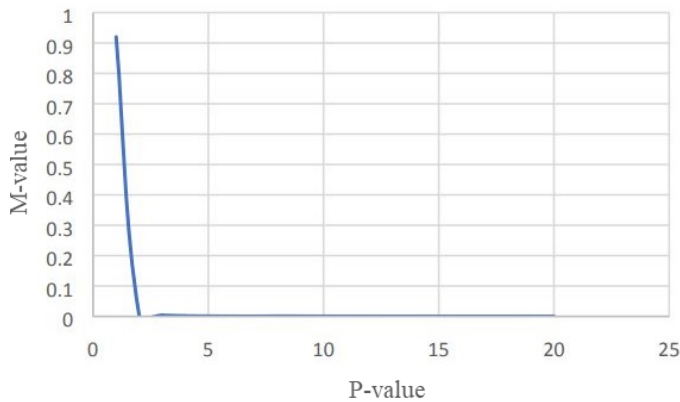


Fig. 3. Graph p vs. m.

Consider the Bloom Filter in which P and Q are two input elements. In Bloom Filter, k represents the number of hash functions of a key or element. In Figure 2, P and Q occupy 3 cells if  $k = 3$ . P is hashed into three different places in the array. In some Bloom filters, fingerprints are stored, while in others, only '1' or '0' are stored. Also, Q is hashed three times in the array. A positive answer of the Bloom Filter requires those three positions to be set (true) when checking the membership of P or Q.

#### F. Overall Work

In situations where non-symmetrical methods such as digital signatures are barely suitable, message authentication codes are commonly implemented in end-to-end authentication. They are most commonly provided in HMAC format. This step-by-step authentication is greatly relevant to broadcast authentication as a privacy enhancer. Hence, WSNs in cellular networks are mostly supplied with this system. Here this system mainly aims to send information to a particular target. Broadcast authentication is very good for securing hop transactions, as every cellular network depends on the broadcast. Delivers a method of hop-by-hop authentication that greatly assures the detection of any imparted false messages by the sink node, where ignoring nodes never crosses the limit. The target should contact adjustment nodes to connect their message authentication codes. Borgia [25] provides a key chain method for LHAP [26], where another keychain is generated to evaluate information from adjacent nodes. Here, consistency

assurance among neighbors is essential. In the case of HEAP [27], every node gets message authentication codes connected, which would increase the band in turn. It is compulsory when ALPHA [28] is concerned with good relations between the source and the target. Even though Curtain [29] becomes the first want to use the bloom filter, authentication of broadcast from the origin is the authors' concern. We have successfully connected the message authentication codes of neighbor nodes with the help of Bloom filters, which can be as efficient as theirs.

$$k = \text{round}((m/n) * \log(2)) \quad (1)$$

$$m = \text{ceil}((n * \log(p)) / \log(1 / \text{pow}(2, \log(2)))) \quad (2)$$

$$p = \text{pow}(1 - \exp(-k / (m/n)), k) \quad (3)$$

$$n = \text{ceil}(m / \log(1 - \exp(\log(p)/k))) \quad (4)$$

#### IV. BASIC DESIGN OF THE MECHANISM

All the route nodes that lie in the band are assumed to be friends because the source has contributed its authentication key with all those nodes. The information M is shared through the routing path  $R_T = \{R_1, R_2, \dots, R_n\}$  to the target node T by the source S. Initially, a fundamental system with an issue is presented, and then we propose a further greater creation is shown inside the subsequent phase. To simplify this, we pass over by using what manner the networks come to a selection and provide routing paths.

##### A. Basic Technique

The construction of the MAC manifest using the Bloom filter is considered the entire working system for step-by-step recognition. En-route authentication with a map is the output obtained by applying the above technique. While other methods use a particular receiver to evaluate the genuineness of information en route authentication, the bitmap uses one or more receivers. En-route authentication code is much smaller than the message authentication code. Thus, EAB utilizes a lower amount of bandwidth for transmitting. Hence it becomes fitter to travel through the routing path. When an en-route authentication code is employed, every router in between is insisted to allow only false information of less liability. Hence, it becomes compulsory for the foe to breach an authorized EAB to impart a false message to a node where each router allows only less liable false messages.

##### B. Protocol Description

1) *Source node*: Firstly, the validation parameters are determined by the source node based on its privacy needs. There begins the multi-hop routing along S, R, and T. The source node initially creates a signature that is responsible for the start of end-to-end authentication. To generate the signature, either message authentication uses a public key code. Only the source node has the message authentication code key. There is no threat to the security by the adversary since M is in the encrypted state. Here the message by S is  $M' \leftarrow \{S | T | M | \text{sig}\}$ .

Every route  $H_i$  gets validated by  $S$  with the help of an en-route authentication bitmap. Using this source node pare wise key, the message authentication code is generated by Eq. 5.

$$EAB[x] = \begin{cases} 1 & \exists_i \exists_j, x = H_j(MAC(KS, R_i, M')) \\ 0 & otherwise \end{cases} \quad (5)$$

When including  $H_i$ , the message authentication code is interlinked with the  $m$ -bit filter; hence, the en-route authentication code is generated when the particular message, as well as its path, is concerned. At last, the initial root node receives the combination of information and its en-route authentication code. Algorithm 1 explains the source node procedure.

2) *Destination node*: The best authentication method for sigS or TM is to authenticate  $M$  at the destination.

3) *Route node*:  $R_i$  gets a message with the EAB BFS,  $R(M')$  for each routing node, which makes certain of the following situations. The number of marked bits in EAB falls within the variety according to Eq. 6.

$$\begin{aligned} & [[mp - \delta], [mp + \delta]] \quad (6) \\ & \delta = \sqrt{mp(1-p)} \\ & p = 1 - (1 - 1/m) k_n \end{aligned}$$

We will make clear the effect of  $\delta$  in the element deferred. If each situation is contented,  $R_i$  will course the message at the side of the filter out to the subsequent router  $R_{i+1}$ . Afterward, it separates the message and reports some false facts. Algorithm 2 describes the procedure in detail.

For instance, as shown in Fig. 4, the source node shapes an augmented Bloom filter of size  $m = 32$  for the diffusion over  $N = 8$  path nodes. Every MAC ( $KS, R_i, M'$ ) maps to  $k = 3$  marked bits in EAB. The ultimate EAB is the bit-wise AND of all plotting. On routine, the EAB will incorporate about 15 marked bits. According to Fig. 5 and 6, the mapped bits in EAB are segmented by MAC.

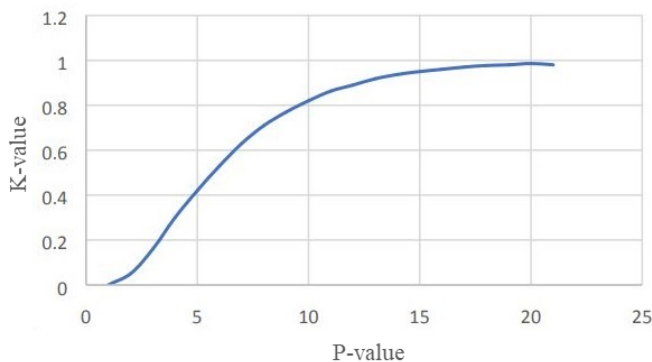


Fig. 4. Graph p vs. k.

### C. Accomplishment of Hash Set

Generally, it is not necessary for  $S$  to implement a new hash set  $h$  to work on every message authentication code. In case cryptographic hash functions are used in message authentication code, the marked bits are automatically defined by the output gained. When the image is fragmented to 160/5, that is 32 pieces. The initial three pieces are selected by  $S$  as a reference to the market bits in the En authentication code. It greatly adds an advantage when the cost is concerned, where the root node must implement only one hash function to validate the information.

### D. Security Analysis

As the message authentication codes are directly involved in the development of en-route authentication code, the MAC keys are confidential; the breach system mainly aims to break through en-route authentication, aiming its false messages, where breaking in maximum probable route nodes is planned. Here, en-route authentication consists of hash functions generated by the message authentication code of various route nodes, in which the attacker has individual distributions. So, ignoring any node on the way could cause the attacker's intentions to fail.

1) *Probability of bypass*: With the usage of a 32-bit en-route authentication bitmap and  $k=3$  hash by the source node, the attack of the adversary reduces up to 81 percent before the completion of two steps. Security increases as the size of the en-route authentication bit increase, but the bandwidth also increases. This attack probability further decreases when 64-bit Android authentication bitmap and  $k=6$  hash function. The adversary must inject a maximum of 1400 information to breach at least the third step node even with half chances. The probability is given by Eq. 7.

$$P_{bypass} = (1 - (\frac{1}{m})^{nk})^t \quad (7)$$

2) *The parameter k*: To calibrate the bloom filter, the amount of hash function is to be calibrated. If  $k = \ln 2m/n$ , the en-route authentication bitmap would contain the maximum false positive.

The parameter  $\delta$ : The attacker can produce fewer Mark bits by optimizing this parameter. This makes the en-route authentication bitmap to be a binomial distribution. It is not necessary e for the mark bit to be  $mp$  always. The common number of marked bits can be between  $\delta = 2\sqrt{mp(1-p)}$  whose chances is higher than 90%. It would be very easy for the attacker to break all the barriers easily without this parameter. There are some disadvantages. The attacker sets the largest possible number of Marked bits instead of the optimum amount to increase the winning chances. In some cases,  $S$  would create an en-route authentication map.

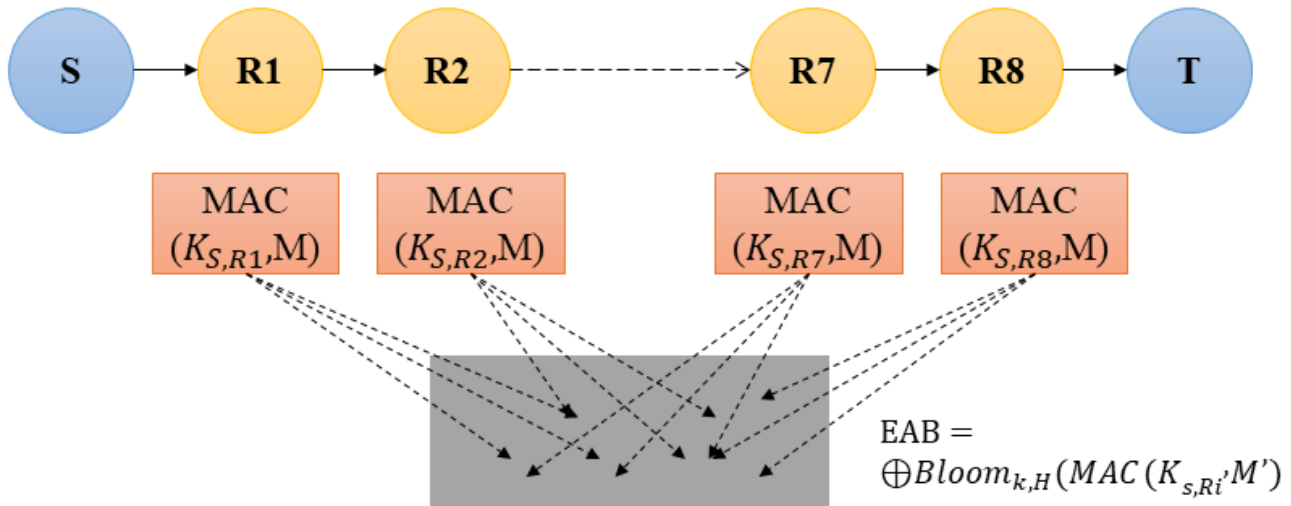


Fig. 5. En-route authentication bitmap.

---

**Algorithm 1.** Source node procedure

---

```

compute M' ← {S | T | M | sigs,T(M)}
clean EAB
for i=1 to n do
    Bitmap  $\xleftarrow{\text{bloom}} H, MAC(K_{S,R_i}, M')$ 
    EAB ⊕ bitmapi
end for
Send out { M'|EAB }

```

---

**Algorithm 2.** Route node R<sub>i</sub> procedure

---

```

Receive packet { M'|EAB } from router Ri-1
If (⌊mp - δ⌋ ≤ count(EAB) ≤ ⌊mp + δ⌋) then
    for j = 1 to k do
        compute x'_j ← H_j(MACS,R_i(M'))
        if (EAB[x'_j] = 0) then
            drop the packet
        end if
    end for
    forward the packet to the next route Ri+1
else
    drop the packet
end if
return

```

---



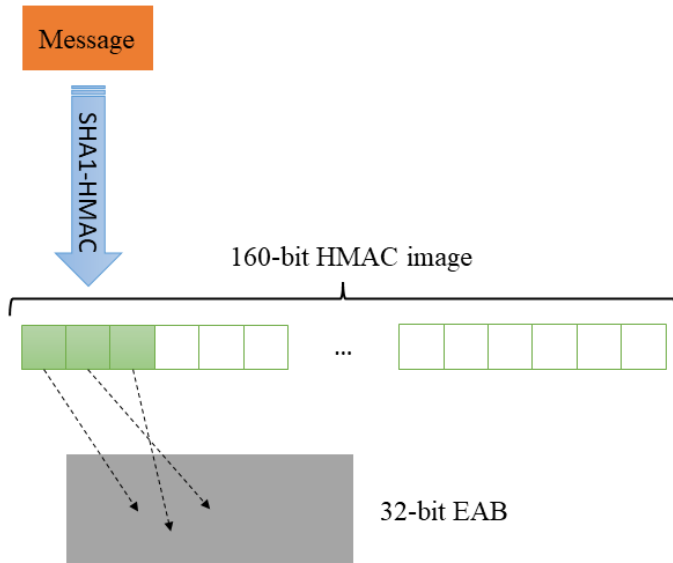


Fig. 6. The segmentation of MAC.

### V. IMPROVEMENT WITH RANDOM PADDING

We have proposed an easy remedy for this problem. This can increase privacy to the next level. The information is as follows when source node  $S$  is supposed to possess two pots.

$$M' \leftarrow \{S|T|M|padding|sig_{S,T}(M|padding)\} \quad (8)$$

Less size of stochastic string that is chosen by  $S$  is said to be a wedding. Hence, the en-route authentication bitmap depends on the number of marked bits. Till the information  $m$  gives an output en-route authentication bitmap with marked bits less than  $\beta$ , the source node sets the padding stochastically until the message is sent. The value of beta is a common and accepted value by each routing node. This helps  $S$  to check different en-route authentication bitmaps for constant information. Hence, it would be easy for the source node to discard insecure en-route authentication bitmaps. However, this checking time method should be reduced if successful results come in series. There is no need for the padding to be larger to succeed in every aspect. In line 2 of algorithm 2,  $S$  changes its characteristics as follows.

$$count(EAB) \leq \beta \quad (9)$$

The approximate number of times of testing the good en-route authentication bitmap as Eq. 8 is as follows.

$$\frac{1}{\sum_{i=1}^{\beta} \binom{m}{i} p^i (1-p)^{m-i}} \quad (9)$$

For example, according to Fig. 7, when  $M$  is 32,  $n$  is 8, and  $k$  is 2, the source node generates an en-route authentication bitmap with marked bits with bits below 13, and this can reduce the injection of false data up to 94.73% in the first and second steps it further gets reduced to 99.72%. When a 4-bit is padded, this probability goes down to  $3.8 \times 10^{21}$ .

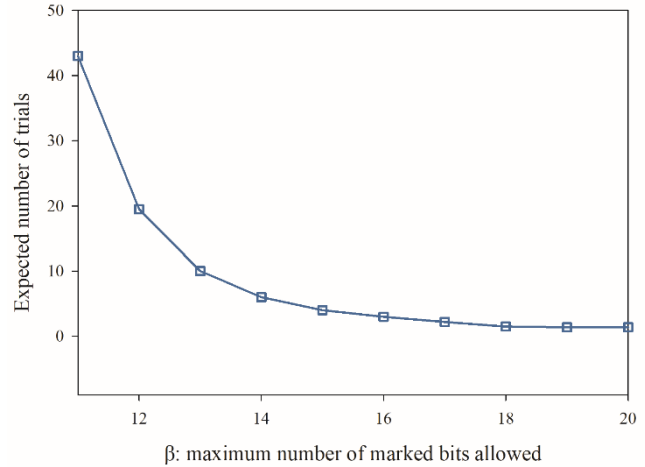
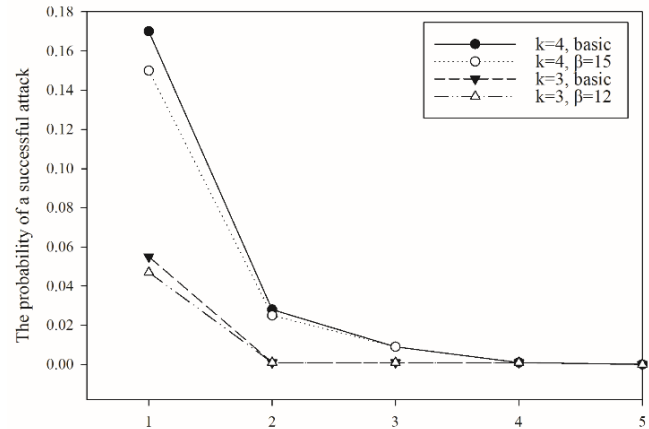


Fig. 7. The expected number of trials to find the desired EAB.

As shown in Fig. 8, for Bloom filters with  $m = 32$  and  $n = 8$ , the optimal  $k$  value should be 3, which will increase the probability of passing through route nodes. The well-selected EAB of  $\beta = 15$  under  $k = 4$  may have superior performance compared to smaller  $\beta$  under smaller  $k$ . Nevertheless, more padding tests or even greater padding will be needed if the number of hashes increases. According to Eq. 10, the parameters depend on the network requirements.



The number of route nodes through which the false injection aims to pass

Fig. 8. The probability of bypass in fine-tuned EAB.

$$\frac{1}{n^S} \leq \left(\frac{\beta}{m}\right)^k \leq P_b \quad (10)$$

### VI. CONCLUSION

This paper proposed a lightweight answer for step-by-step authentication in IoT. Low computational cost, high accuracy, and low communication overhead characterize our shape. Authentication is performed by combining Augmented Bloom filters with bitmaps from en-route authentication. The Augmented bloom filter utilizes a single hash function named Message Digest (MD5). Its main intention is to thoroughly remove the false negative rate and highly lessen the false positive rate. Single hash function usage results from more

false positive rates, but we impuse the MD5 here. We split the input into three and applied the MD5 hash function to each of them separately to collect three different key values. This idea works great on it and successfully results in a less false positive rate. The most important goal of EAB is to categorize false records inside the first limited steps as a substitute for dropping them at the beginning sign. Thus, we admit an adequate false-negative rate adjoining the adversary, but the false-negative rates quickly come together as the next count increases. This interchange significantly diminishes the communique overhead received through authentication. We declare that such remedy can be retained similarly to light-weight authentication outlines in cooperative environments.

#### REFERENCES

- [1] Mousavi, S.K., et al., Security of internet of things based on cryptographic algorithms: a survey. *Wireless Networks*, 2021. 27(2): p. 1515-1555.
- [2] Ataie, I., et al. D 2 FO: Distributed Dynamic Offloading Mechanism for Time-Sensitive Tasks in Fog-Cloud IoT-based Systems. in *2022 IEEE International Performance, Computing, and Communications Conference (IPCCC)*. 2022. IEEE.
- [3] Seyfollahi, A., T. Taami, and A. Ghaffari, Towards developing a machine learning-metaheuristic-enhanced energy-sensitive routing framework for the internet of things. *Microprocessors and Microsystems*, 2023. 96: p. 104747.
- [4] Mehbodniya, A., et al., Energy-Aware Routing Protocol with Fuzzy Logic in Industrial Internet of Things with Blockchain Technology. *Wireless Communications and Mobile Computing*, 2022. 2022.
- [5] Sellami, B., et al., Energy-aware task scheduling and offloading using deep reinforcement learning in SDN-enabled IoT network. *Computer Networks*, 2022. 210: p. 108957.
- [6] Haghshenas, S.H., M.A. Hasnat, and M. Naeini, A Temporal Graph Neural Network for Cyber Attack Detection and Localization in Smart Grids. *arXiv preprint arXiv:2212.03390*, 2022.
- [7] Mishra, S. and A.K. Tyagi, The role of machine learning techniques in internet of things-based cloud applications, in *Artificial Intelligence-based Internet of Things Systems*. 2022, Springer. p. 105-135.
- [8] Cauteruccio, F., et al., A framework for anomaly detection and classification in Multiple IoT scenarios. *Future Generation Computer Systems*, 2021. 114: p. 322-335.
- [9] Ferrag, M.A. and L. Shu, The performance evaluation of blockchain-based security and privacy systems for the Internet of Things: A tutorial. *IEEE Internet of Things Journal*, 2021. 8(24): p. 17236-17260.
- [10] Ren, P., et al., IPSadas: identity - privacy - aware secure and anonymous data aggregation scheme. *International Journal of Intelligent Systems*, 2022. 37(8): p. 5290-5324.
- [11] Shah, P. and T. Kasbe, A review on specification evaluation of broadcasting routing protocols in VANET. *Computer Science Review*, 2021. 41: p. 100418.
- [12] Montanari, A.N., et al., Functional observability and target state estimation in large-scale networks. *Proceedings of the National Academy of Sciences*, 2022. 119(1): p. e2113750119.
- [13] Pourghebleh, B., et al., A roadmap towards energy - efficient data fusion methods in the Internet of Things. *Concurrency and Computation: Practice and Experience*, 2022: p. e6959.
- [14] Mohseni, M., F. Amirghafouri, and B. Pourghebleh, CEDAR: A cluster-based energy-aware data aggregation routing protocol in the internet of things using capuchin search algorithm and fuzzy logic. *Peer-to-Peer Networking and Applications*, 2022: p. 1-21.
- [15] Saeidi, S.A., et al. A novel neuromorphic processors realization of spiking deep reinforcement learning for portfolio management. in *2022 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. 2022. IEEE.
- [16] Zheng, Y., et al., PUF-based Mutual Authentication and Key Exchange Protocol for Peer-to-Peer IoT Applications. *IEEE Transactions on Dependable and Secure Computing*, 2022.
- [17] Chen, X., et al. A blockchain based access authentication scheme of energy internet. in *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*. 2018. IEEE.
- [18] Lawrence, T., et al., A computationally efficient HMAC-based authentication scheme for network coding. *Telecommunication Systems*, 2022. 79(1): p. 47-69.
- [19] Nianmin, Y., M. Haifeng, and H. Yong, A method for memory integrity authentication based on bloom filter. *Journal of Algorithms & Computational Technology*, 2014. 8(3): p. 267-286.
- [20] Saravanan, K. and A. Senthilkumar, Security enhancement in distributed networks using link-based mapping scheme for network intrusion detection with enhanced Bloom filter. *Wireless Personal Communications*, 2015. 84(2): p. 821-839.
- [21] Wang, J., et al. An authentication key agreement scheme for heterogeneous sensor network based on improved counting bloom filter. in *2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)*. 2015. IEEE.
- [22] Malhi, A. and S. Batra, Privacy-preserving authentication framework using bloom filter for secure vehicular communications. *International Journal of Information Security*, 2016. 15(4): p. 433-453.
- [23] Mbarek, B., N. Sahli, and N. Jabeur. BFAN: A Bloom Filter-Based Authentication in Wireless Sensor Networks. in *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*. 2018. IEEE.
- [24] Sateesan, A., et al. Novel Bloom filter algorithms and architectures for ultra-high-speed network security applications. in *2020 23rd Euromicro Conference on Digital System Design (DSD)*. 2020. IEEE.
- [25] Borgia, E., The Internet of Things vision: Key features, applications and open issues. *Computer Communications*, 2014. 54: p. 1-31.
- [26] Zhu, S., et al. LHAP: a lightweight hop-by-hop authentication protocol for ad-hoc networks. in *23rd International Conference on Distributed Computing Systems Workshops*, 2003. Proceedings. 2003. IEEE.
- [27] Akbani, R., T. Korkmaz, and G. Raju. HEAP: Hop-by-hop efficient authentication protocol for mobile ad-hoc networks. in *Proceedings of the 2007 spring simulaiton multicference-Volume 1*. 2007.
- [28] Heer, T., et al. Alpha: an adaptive and lightweight protocol for hop-by-hop authentication. in *Proceedings of the 2008 ACM CoNEXT Conference*. 2008.
- [29] Chen, Y.-S., et al. Broadcast authentication in sensor networks using compressed bloom filters. in *International Conference on Distributed Computing in Sensor Systems*. 2008. Springer.

# Generalized Epileptic Seizure Prediction using Machine Learning Method

Zarqa Altaf<sup>1\*</sup>, Mukhtiar Ali Unar<sup>2</sup>, Sanam Narejo<sup>3</sup>, Muhammad Ahmed Zaki<sup>4</sup>, Naseer-u-Din<sup>5</sup>  
Mehran University of Engineering and Technology, Jamshoro, Sindh, Pakistan<sup>1, 2, 3, 5</sup>  
NED University of Engineering and Technology, Karachi, Sindh, Pakistan<sup>4</sup>

**Abstract**—In recent years, the electroencephalography (EEG) signal identification of epileptic seizures has developed into a routine procedure to determine epilepsy. Since physically identifying epileptic seizures by expert neurologists becomes a labor-intensive, time-consuming procedure that also produces several errors. Thus, efficient, and computerized detection of epileptic seizures is required. The disordered brain function that causes epileptic seizures can have an impact on a patient's condition. Epileptic seizures can be prevented by medicine with great success if they are predicted before they start. Electroencephalogram (EEG) signals are utilized to predict epileptic seizures by using machine learning algorithms and complex computational methodologies. Furthermore, two significant challenges that affect both expectancy time and genuine positive forecast rate are feature extraction from EEG signals and noise removal from EEG signals. As a result, we suggest a model that offers trustworthy preprocessing and feature extraction techniques. To automatically identify epileptic seizures, a variety of ensemble learning-based classifiers were utilized to extract frequency-based features from the EEG signal. Our algorithm offers a higher true positive rate and diagnoses epileptic episodes with enough foresight before they begin. On the scalp EEG CHB-MIT dataset on 24 subjects, this suggested framework detects the beginning of the preictal state, the state that occurs before a few minutes of the onset of the detention, resulting in an elevated true positive rate of (91%) than conventional methods and an optimum estimation time of 33 minutes and an average time of prediction is 23 minutes and 36 seconds. Depending on the experimental findings' The maximum accuracy, sensitivity, and specificity rates in this research were 91%, 98%, and 84%.

**Keywords**—Epilepsy; electroencephalogram; artificial intelligence; machine learning; CHB-MIT

## I. INTRODUCTION

A set of neurological illnesses known as epilepsy can afflict people of any age and are defined by a persistent propensity to cause repeated seizures. The progressive neurobiological process known as "epileptogenic" causes epilepsy [1]. The aberrant synchronized electrical activity of brain neurons is the primary cause of epilepsy, a persistent, non-communicable condition [2, 3]. The oldest and most prevalent neurological condition in the globe is epilepsy [4, 5]. Epilepsy is the third most prevalent neurological condition in the world, affecting 50 million individuals worldwide, based on a World Health Organization (WHO) study from June 2019 [6]-[10]. An abnormality of the brain characterized by recurrent seizures is called Epilepsy. Typically, a seizure is described as a sudden (abrupt) shift in behavior because of an

abnormal disturbance in the electrical activity of the human brain [11]. Some minute electrical impulses are continuously produced by the brain resulting in a consistent pattern. Neurotransmitters are the chemical signals which carry electrical signals along with neurons, and neural networks in the brain and throughout the entire body [12].

Fig. 1 illustrates how epilepsy causes the brain's electrical cycles to become unbalanced and cause recurring seizures. Individuals having seizures must face synchronized electrical energy bursts that may alter their cognition, movements, or perceptions and disturb the regular brain electrical sequence for a period. The main symptoms of epilepsy are varied and complex due to variations in the beginning location and method of propagation of aberrant electrical activity in the brain [13]. Recurrent seizures can have a long-lasting severe impact on a patient's psychological and cognitive abilities and pose a serious risk to their lives [14]. Investigation into the treatment and diagnosis of epilepsy, therefore, has huge therapeutic implications.

Epileptic seizures can be prevented by medication if they are predicted early, giving ample time before they happen. Four distinct states occur during epileptic seizures. The first state that emerges before the beginning of the seizure is the prodromal (pre-ictal) state, the second state i.e., the ictal state, starts with the exact beginning of the seizure and completes leaving a threat, after the ending of the ictal state comes the third state i.e., postictal state, and last is an interictal state, that begins after the postictal state of the first seizure and is finished before the begin of the preictal state of the subsequent seizure. The various input conditions for three distinct channels are depicted in Fig. 2. Additionally, the onset of the preictal state can be used to anticipate seizures [15].

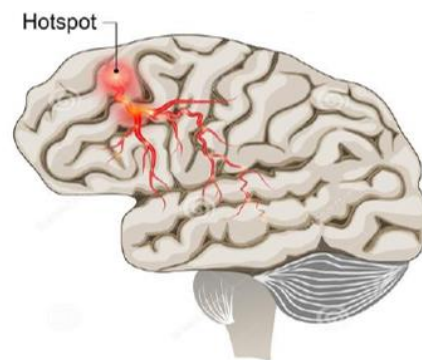


Fig. 1. Epilepsy hotspot.

\*Corresponding Author.

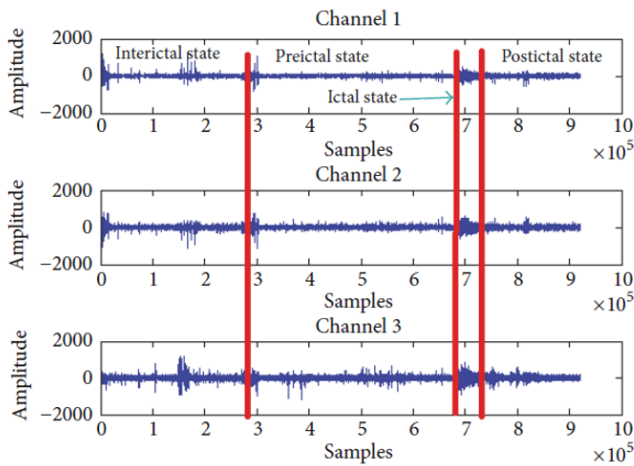


Fig. 2. Different input states of epileptic seizure [15].

The remainder of the essay has been structured as follows: The complaints are covered in Section II, the background of epilepsy seizures is covered in Section III, and the proposed technique is covered in Section IV. The experimental results are reported in Section V. Section VI brings the essay to a close and discusses unfinished business.

## II. TYPES AND SYMPTOMS OF EPILEPTIC SEIZURE

Neurologically epilepsy is characterized by abnormal activity of the brain that results in seizures resulting in strange behavior, emotional sensations, and most of the time total loss of conscious [16]. When a person experiences at least 2 seizures that are not related to another established medical problem, such as opiate withdrawal or exceptionally low blood sugar, an epilepsy diagnosis is typically made [17]. That part of the brain from which the seizure frequently originates in early phases causes disturbance in functions of the affected part. The right side of the body is governed by the left half of the brain, while the left side of the body is governed by the right half of the brain. Typically, Doctors determine seizure as either generalized or focal depending on where and how the abnormal activity of the brain starts [18]. Focal seizures are caused by the aberrant activity of the brain in a specific part of the brain, while Generalized seizures appear to be involved in the entire brain [19]. Neuro-experts have divided seizures into two main groups, partial and generalized, depending on the signs, as depicted in Fig. 3 [20, 21]. The symptoms of a partial seizure, which are mostly brought on by damage to the cerebral hemisphere, can be utilized to define it. Additionally, there are two basic categories of partial seizures: simple-partial and complex-partial. In simple-partial, the person appears cognizant and can typically speak, whereas, in complex-partial, patients behave erratically, become disoriented, and frequently mumble and chew. A generalized seizure comprises two main components as well. While definitive seizures are challenging to detect because they lack motor signals, non-conclusive seizures can be identified by their clear motor symptoms. The person is unable to move or say anything other than to gaze [22, 23].

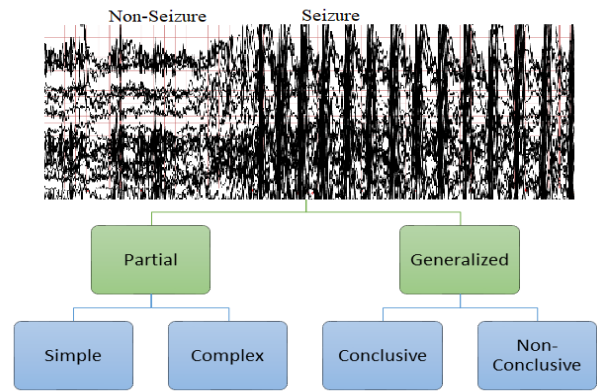


Fig. 3. Types of epileptic seizures.

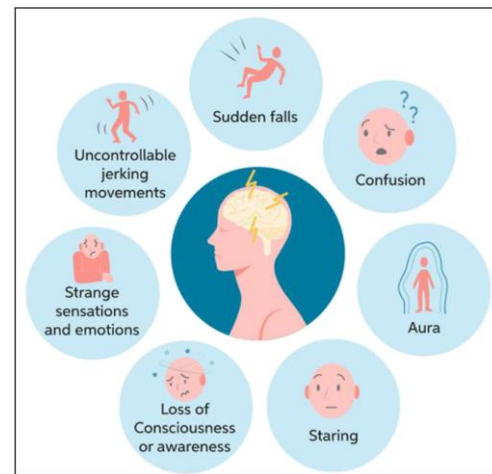


Fig. 4. Symptoms of epileptic seizure [23].

Fig. 4 illustrates the wide range of seizure signs. Throughout a seizure, some patients just stare aimlessly for a specific period, while others continuously jerk their limbs or legs. One seizure may not necessarily indicate epilepsy. For an epileptic classification, at least 2 unprovoked seizures (seizures caused by unknown reasons) must be occurred within 24 hours away [24, 25]. Any brain-coordinated process can be disturbed by seizures since aberrant brain activity causes Epilepsy. Some specific symptoms determine the epilepsy type. Some of the below-mentioned sensations will be realized from time to time while others become consistent. Most of the time, an individual having epilepsy experiences the same type of seizure every other time. Symbols and Seizure indications may contain [26]:

- Brief uncertainty (confusion).
- Steady Eye spell.
- Rigid Body Movement.
- Uncontrollable spasmodic motion.
- Unawareness and incognizance.
- Spiritual indications such as fear or nervousness.

### III. RELATED WORK

Early studies on Epilepsy prediction were conducted in the 1970s utilizing feature extraction methods that were linear [27]. Because of the non-linear character of EEG signals in the 1980s, researchers were able to apply these approaches for feature extraction thanks to the advent of non-linear methods [28, 29]. The utilization of the pre-ictal phase for epilepsy identification was also implemented in this decade with the diagnosis of the EEG patterns associated with epilepsy, including preictal, ictal, and interictal patterns. Salant et al. conducted early ES prediction almost 6 seconds before the seizure began in 1998 [30], and Drogenlen et al. 2003 expanded on this work [31]. They employed a feature called Kolmogorov entropy to forecast epilepsy 2 – 40 minutes before it began. The very first worldwide session on epilepsy forecasting took place in 2002, and several epilepsy facilities contributed a database of multi-day EEG recordings. Eventually, this database was the subject of other investigations [32]. Mormann et al. discovered in 2003 that the periodic synchronization of various EEG channels diminishes before seizure onset [33] using this theory that the hyper-synchronous discharge of the brain's neurons causes ES. Research studies on substantial EEG data have cast doubt on the accuracy of metrics computed in the past century during the first 10 years of the ongoing century. Some researchers discovered that these findings belonging to past studies were based on a limited number of carefully chosen data that could not be replicated on a large amount of previously unreported data. In worldwide workshops held on the subject, it was determined to hold contests on seizure prediction. These contests were created to make it easier to compare the effectiveness of algorithms that had been trained on the same dataset [34, 35]. The International Workshop on Seizure Prediction 3 (IWSP3) and the International Workshop on Seizure Prediction 4 (IWSP4) collaborated on the inaugural seizure prediction competition, which took place in 2007. The participants in both events received continuous iEEG recordings from 3 epilepsy patients. The algorithms' results obtained, however, fell short of expectations.

The 2014 American Epilepsy Society Seizure Forecasting Trial used long-term iEEG recordings of epileptic canines as well as short-term human iEEG containing 942 seizures acquired over more than 500 days. The same training and testing data, lasting 10 minutes, was given to each contestant. An evaluation metric for effectiveness was the Area Under the Curve (AUC). Another competition by Melbourne University with a similar format comprised long-term iEEG recording with 1139 seizures [36]. Any algorithm estimating the fundamental properties of EEG signals for epilepsy predictions or machine learning algorithms based on these basic properties was eligible for the competition. In any scenario, we are still unsure of the ideal characteristics or techniques. People entered algorithms that were excessively complex in the competitions. Therefore, it is challenging to determine which attribute or ML method was better. A novel solution presented by Maturana et al. [37] may be effective for a variety of patients. They determined that the crucial slowing of neural activity served as an ES prediction indication. Fig. 5 shows a timeline for the evolution of EEG data measurements. Readers who are interested in learning more about the

background of these advances should consult [38] for additional details. Fig. 6 from Natu et al. [39] discussion on the development of technology for epileptic seizure detection.

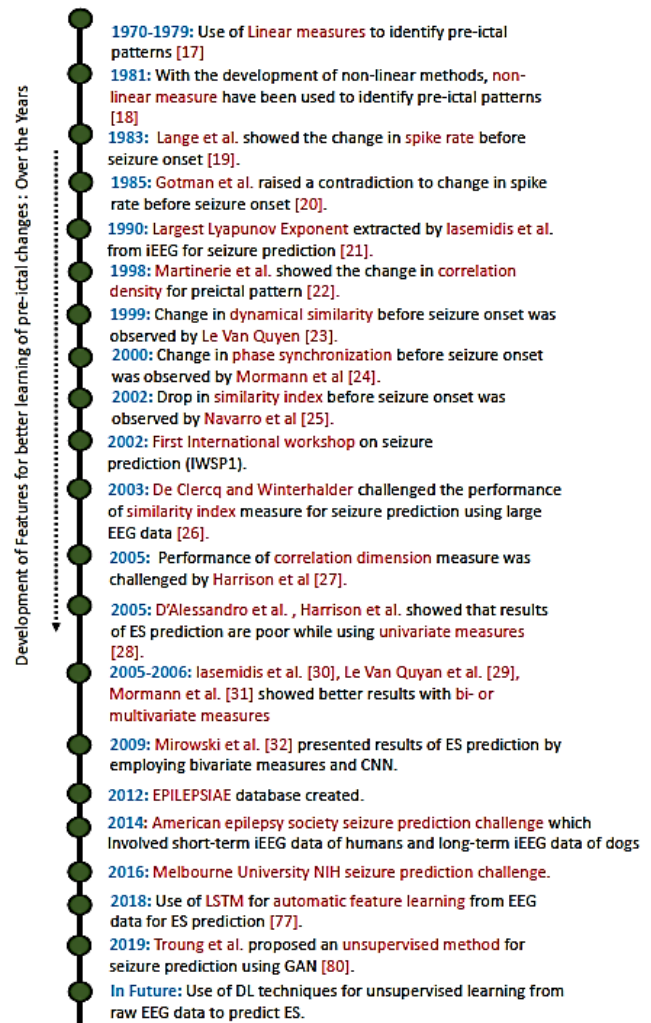


Fig. 5. History of epileptic seizure prediction [38].

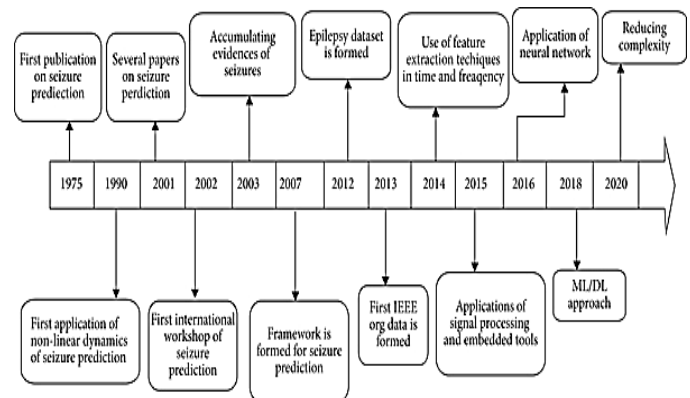


Fig. 6. Machine learning for generalized epileptic seizure prediction.

#### IV. MACHINE LEARNING FOR GENERALIZED EPILEPTIC SEIZURE PREDICTION

Since the turn of the century, scientists have been trying to get past the difficulties in diagnosing and predicting epilepsy. The initial emphasis of the ES forecast study was mainly on the evaluation of EEG recordings because EEG data are an important source to observe brain function before, throughout, and after epilepsy. Eye rotations, blinks, heart signals, and muscular noise contaminate EEG signals. To lessen the impact of these numerous sources of interference and distortions, a variety of filtration and noise reduction techniques are employed [40]. Substantial features are required for developing Machine learning models for the classification and identification of interictal and pre-ictal phases once artifacts have been removed. Fig. 7 illustrates the traditional Machine learning approach for epilepsy forecasting and emphasizes the key distinction between the application of Machine learning and Deep Learning methods.

##### A. Signal Processing

One important step in the analysis of raw biological signals is the identification of noise and artifacts. Filtering of these artifacts is required to lessen their impact on feature extraction. For filtering, a variety of methods have been used, including many filters such as Wavelet, Band-Pass, Finite Impulse response, and adaptive filters. Additionally, such processing is done to make the data standardized so that it may be compared to the records of other patients.

##### B. Feature Extraction and Collection

Reliable features are a requirement for all prediction models. These features can be divided into unilateral (steps undertaken on each EEG channel independently) and multimodal (measurements taken on two or more EEG channels) categories according to the quantity of EEG channels. Numerous techniques recommended in the literature were used to do the EEG study. As shown in Fig. 8, these methods were widely divided into 4 categories: frequency domain, time domain, nonlinear approaches, and time-frequency domain.

##### C. Classification

Artificial neural networks (ANN), fuzzy logic, k-means clustering, support vector machines (SVM), and decision trees are used to ensure the identification of epileptic seizures from provided EEG data. Most of the time, feature values with thresholds are used to draw inferences.

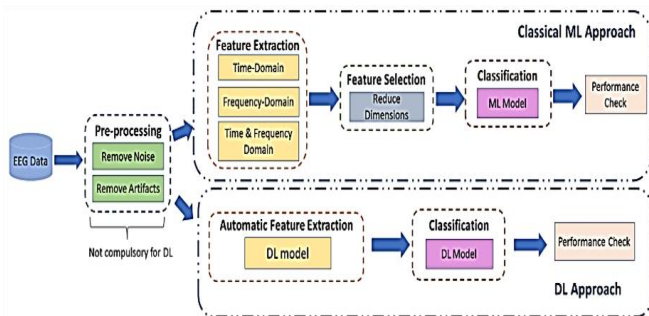


Fig. 7. Different input states of epileptic seizure [15].

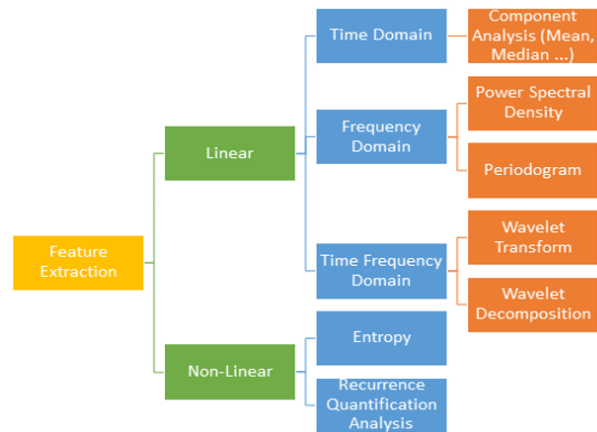


Fig. 8. Classification of feature extraction approaches [42].

#### V. GENERALIZED EPILEPSY PREDICTION USING ML METHOD EXPERIMENTAL DETAILS

The primary goal of this research is to use computer vision algorithms to classify EEG signals as epileptic signals (pre-ictal phase) or non-epileptic signals for the diagnosis of epilepsy. The signals in this are centered on ictal release for epileptic signals, whereas non-epileptic signals are consisting of both normal and pathological inter-ictal discharges for non-epileptic signals. The technique utilized to accomplish this is as follows:

- 1) EEG signal normalization and signal extraction.
- 2) To generate a feature collection, extract statistical features.
- 3) Apply wavelet decomposition to the signal to break it down.
- 4) To reduce the runtime, using k-means clustering for reducing the number of features in the feature set.
- 5) Training of the Support Vector Machine using the condensed feature set.
- 6) On a test data set, compare how well the SVM is based on the entire and modified feature set performed in separating epileptic from non-epileptic signals.

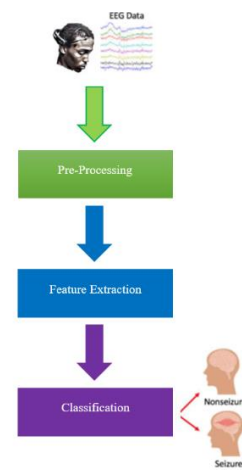


Fig. 9. Machine learning proposed model.

Fig. 9 displays the suggested method's block diagram. The gathering of EEG datasets, the pre-processing of these signals, feature extraction, and classification are the four primary stages of this approach. Below is a detailed explanation of these actions.

**D. EEG Dataset Collection**

The CHB-MIT database, which contains EEG recordings, was used in this investigation. All signals were captured and made accessible to the public by Boston Children's Hospital. Many recordings last an hour, but others go on for two or four hours. 24 portions of an EEG recording are separated and recorded in the EDF database schema. An EEG recording is represented by each EDF file. The CHB-MIT dataset signals include 686 EEG recordings from 23 people ranging in age from 1.5 to 17 years old. Each participant is represented by several EEG signals from various channels, and the dataset's sample frequency is 256 Hz. The Chb01 (1st subject) and Chb21 (second subject) in this database are the same individuals who were enrolled over 1.5 years. Information from the CHB-MIT dataset is shown in Table I.

Fig. 10 depicts a sample EEG signal from a person experiencing an epileptic seizure. In this image, the Seizure time window is indicated by red lines. Instances of (a) non-seizure records and (b) seizure records from the CHB-MIT database are shown in Fig. 11 [41]. Dataset information from EEGLAB executing in MATLAB is shown in Fig. 12.

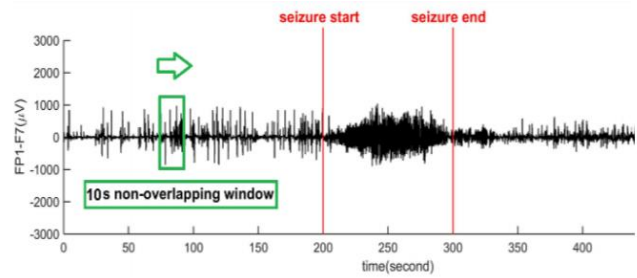


Fig. 10. EEG signal with epilepsy seizure [40].

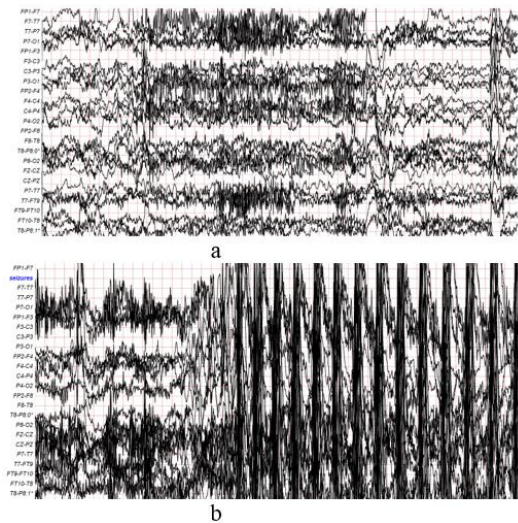


Fig. 11. CHB-MIT database: (a) Non-Seizure; (b) Seizure.

TABLE I. CHB-MIT DATASET

Case No.	Gender	Age	No. of Seizure
1	F	11	7
2	M	9	3
3	M	61	7
4	F	14	4
5	M	45	5
6	M	3	10
7	F	13	3
8	M	76	5
9	F	36	4
10	M	55	7
11	F	6	3
12	F	14	27
13	M	44	10
14	M	4	8
15	F	6	3
16	M	6	6
17	F	5	3
18	M	9	8
19	F	8	4
20	F	27	3
21	M	23	7
22	F	33	16
23	F	7	7
24	M	21	4

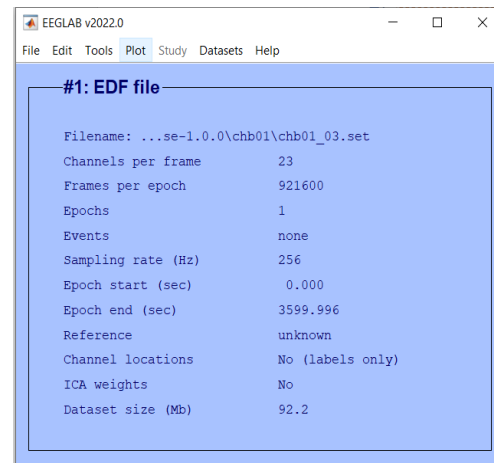


Fig. 12. Dataset details from EEGLAB running in MATLAB.

**E. Pre-Processing of EEG Signals**

Preprocessing is the procedure of transforming raw data into a format that is more suitable for further analysis and interpretable for the user. In the case of EEG data, preprocessing usually refers to removing noise from the data to get closer to the true neural signals.

There are several reasons for preprocessing of EEG data is necessary. First, the signals that are picked up from the scalp are not necessarily an accurate representation of the signals originating from the brain, as the spatial information gets lost.

Secondly, EEG data tends to contain a lot of noise which can obscure weaker EEG signals. Artifacts such as blinking, or muscle movement can contaminate the data and distort the picture. Finally, we want to separate the relevant neural signals from random neural activity that occurs during EEG recordings. Fig. 13 represents the preprocessing pipeline that is followed in this research.

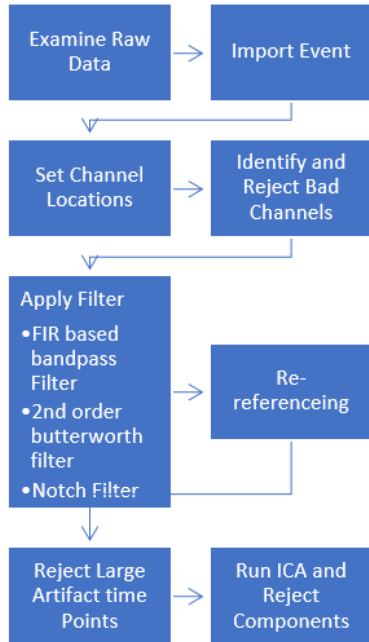


Fig. 13. Preprocessing pipeline.

After following the preprocessing pipeline, we reduce the 23 channels into 8 channels which are:

- 1) FP1-F7
- 2) P3-O1
- 3) P4-O2
- 4) FP2-F8
- 5) P8-O2
- 6) FZ-CZ
- 7) CZ-PZ
- 8) P7-T7

The results of these 8 channels are shown in Fig. 14. The dataset with the following setting is shown in Fig. 15. Save the dataset as a '.set' file extension for the next step which is feature extraction.

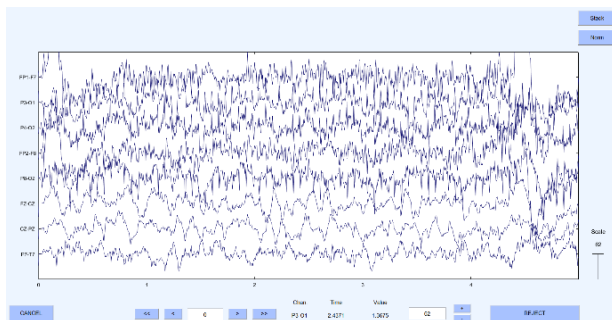


Fig. 14. Visualization of 8-selected channels.

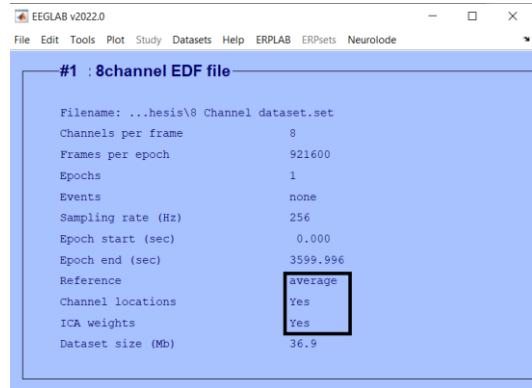


Fig. 15. Data of 8 selected channels.

### F. Feature Extraction

To collect the abstract information required for the classification procedure at this point, feature extraction was used to remove the duplicate information from the EEG signals. When analyzing signals using wavelet transform, it's crucial to choose the right wavelets and the right number of layers of decomposition. The signal's prominent frequency components are used to determine the number of decomposition levels. The amount of decay is selected so that the wavelet coefficients preserve the frequencies necessary for the identification of the signal. The MATLAB software program was used to calculate the wavelet coefficients. In this study, we extract several features including the Fast Fourier transform, wavelet transforms, Mean, and Standard Deviation for alpha, beta, theta, delta, and gamma frequencies as shown in Fig. 16. 12 Features value extracted and saved as '.mat' for normal and epilepsy as shown in Fig. 17.

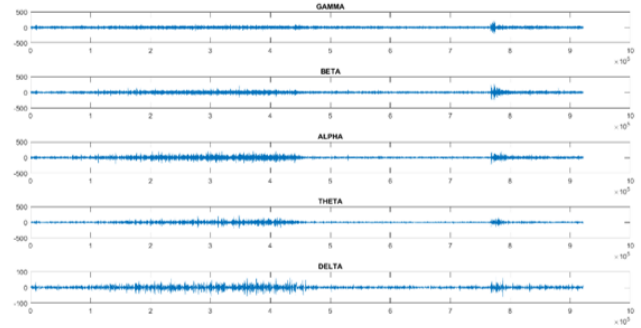


Fig. 16. Alpha, beta, and gamma features.

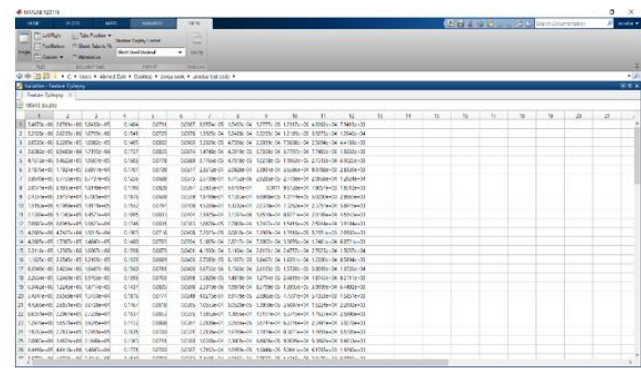


Fig. 17. Feature extraction.



### G. Classification

The goal of the model was to determine the most effective dimensionality reduction method that, when combined with SVM, would provide the maximum degree of sensitivity and validity for gathering statistics as either epileptic or not. In a high-dimensional space, the support vector machine (SVM) creates a hyperplane or series of hyperplanes that can be utilized for classification. SVM has been demonstrated to be a useful supervised model based on a statistical learning tool with high generalization. The principle underlying SVM is the separation of two data sets. This separation can be linear or non-linear. In the case of linear separation, SVM uses a discriminant hyperplane to distinguish classes. However, in the case of nonlinear separation, SVM uses the kernel function to identify decision boundaries. Compared with that of other supervised algorithms, such as ANNs [42, 43] and KNN, the computational complexity of SVM is low [44]-[46].

In this study, the model of all data of 23 people is used for training each time, and data of the remaining 1 subject is used for the test. We explore multiple training, validation, and testing divisions of the dataset to see the effect on the performance achieved on these subsets. With an increase in training data as compared to testing data, an increase in performance for accuracy and sensitivity is observed. In our experimentation, a train-validation-test ratio of (70%-20%-10%) is followed. This ratio resulted in a total of 50 epochs for training. Fig. 18 shows the Epileptic seizure detection training, validation, and testing of 24 patients.

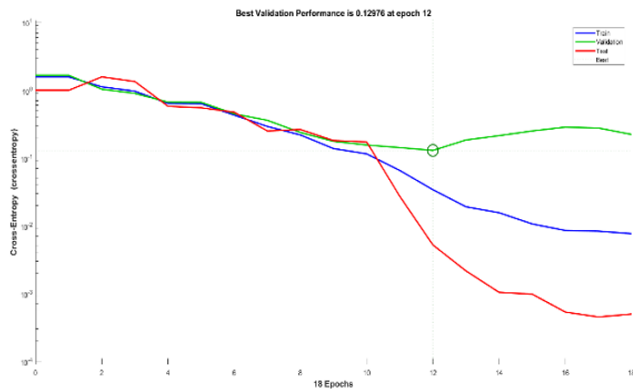


Fig. 18. Epileptic seizure detection training, validation, and testing.

## VI. RESULTS

The clinical employment of ES prediction methods requires a sufficient performance and quality check and different evaluation metrics have been discussed in this section. Our end goal is to classify data into two classes non-seizure and seizure. To measure the performance of the proposed method, a confusion matrix, shown in Fig. 19 is obtained. In this table, (TP) represents true positive (epileptic region predicted as epileptic), TN represents true negative (non-epileptic region predicted as non-epileptic), FP represents false positive (non-epileptic region predicted as epileptic), and FN represents false negative (epileptic region predicted as non-epileptic).

One main challenge in classifying seizure data is the imbalance of the dataset. This comes from the fact that seizures (and so preictal data) do not occur frequently and the size of the interictal class is much larger than the preictal class. This may cause naive classification, which means that the classifier labels all the data as interictal and completely ignores the other class, and still reports a good precision. To avoid this, we propose a few contingency plans. First, we do not rely only on accuracy as the main factor to choose the best classifier. More informative factors can be sensitivity and specificity. Here, accuracy is the correct classification rate, sensitivity is the proportion of the epileptic regions that are correctly classified and specificity is the proportion of the non-epileptic regions that are correctly classified. Sensitivity is defined as the ratio of the total number of true positives (TP) to the sum of the total number of true positives and false negatives (FN). True positive is defined as the detection of a seizure in a segment which is also identified as a seizure segment by experts. Whereas false negatives represent a seizure segment not being classified as so by the algorithm, while the segment is identified as a seizure segment by experts. Specificity is defined as the ratio of the total number of true negatives (TN) to the sum of the total number of true negatives and false positives (FP). True negative is defined as the detection of a non-seizure segment which is also identified as a non-seizure segment by experts. Whereas false positives represent a seizure segment being classified by the algorithm, while the segment is identified to be a non-seizure segment by experts. Accuracy is defined as the ratio of the sum of TP and TN to the sum of TP, TN, FP, and FN. Hence, the higher the value, the better the performance is achieved. Classification results are shown in Table II.

TABLE II. CLASSIFICATION RESULTS

Measure	Value	Derivations
<b>Sensitivity</b>	0.980	$TPR = TP / (TP + FN)$
<b>Specificity</b>	0.840	$SPC = TN / (FP + TN)$
<b>Precision</b>	0.859	$PPV = TP / (TP + FP)$
<b>Negative Predictive Value</b>	0.976	$NPV = TN / (TN + FN)$
<b>False Positive Rate</b>	0.160	$FPR = FP / (FP + TN)$
<b>False Discovery Rate</b>	0.140	$FDR = FP / (FP + TP)$
<b>False Negative Rate</b>	0.020	$FNR = FN / (FN + TP)$
<b>Accuracy</b>	0.910	$ACC = (TP + TN) / (P + N)$
<b>F1 Score</b>	0.915	$F1 = 2TP / (2TP + FP + FN)$

In Fig. 19, the first two diagonal cells show the number and percentage of correct classifications by the trained network. For example, 98 recordings are correctly classified as benign. This corresponds to 49.0% of all 200 recordings. Similarly, 84 cases are correctly classified as malignant. This corresponds to 42.0% of all recordings. 16 of the malignant recordings are incorrectly classified as benign and this corresponds to 8.0% of all 200 recordings in the data. Similarly, 2 of the benign recordings are incorrectly classified as malignant and this corresponds to 1.0% of all data. Out of 114 benign predictions, 86.0% are correct and 14.0% are wrong. Out of 86 malignant predictions, 97.7% are correct and 2.3% are wrong. Out of 100 benign cases, 98.0% are correctly

predicted as benign and 2.0% are predicted as malignant. Out of 100 malignant cases, 84.0% are correctly classified as malignant and 16.0% are classified as benign. Overall, 91.0% of the predictions are correct and 9.0% are wrong. Considering the proposed approach system's specificity (SP), sensitivity (SN), and accuracy (AC) allows for an evaluation of its performance. The proportion of the total number of true positives (TP) to the total number of false negatives and true positives is known as the sensitivity (FN). True positive is the identification of a seizure in a part that is also recognized by professionals as a seizure section. False negatives, on the other hand, refer to a seizure segment that is recognized as a seizure segment by specialists but is not classified as such by the algorithm. Therefore, better performance is obtained as the greater the value. The classification time of SVM is shown in Table III.

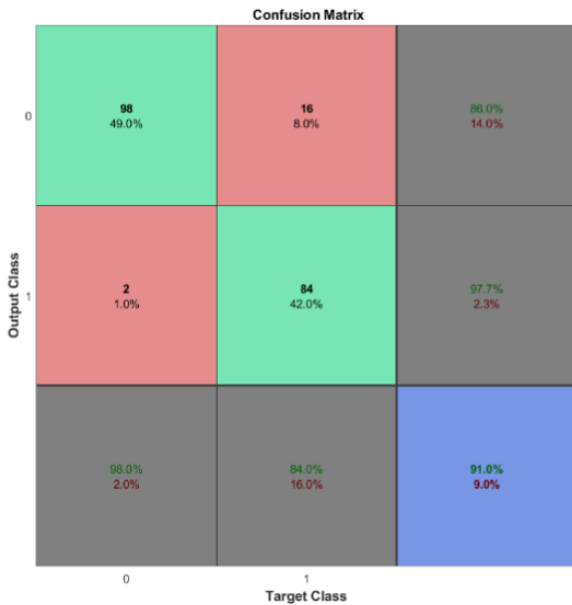


Fig. 19. Confusion matrix.

TABLE III. CLASSIFICATION TIME

Classifier	Training Time (Sec)	Testing Time (Sec)
Support Vector Machine (SVM)	0.150	0.050

## VII. CONCLUSION

The automatic approaches for detecting epileptic seizures have been suggested in this paper. Data from CHB MIT were utilized to detect seizure events. An SVM classifier was used for classification, and maximum accuracy of 90.7% was attained. Training of the classification algorithm was carried out across patients to assess the effectiveness of the suggested method, and the experimental findings were as a result. The maximum accuracy, sensitivity, and specificity rates in this research were 91.0%, 98%, and 84% correspondingly as shown in Fig. 20.

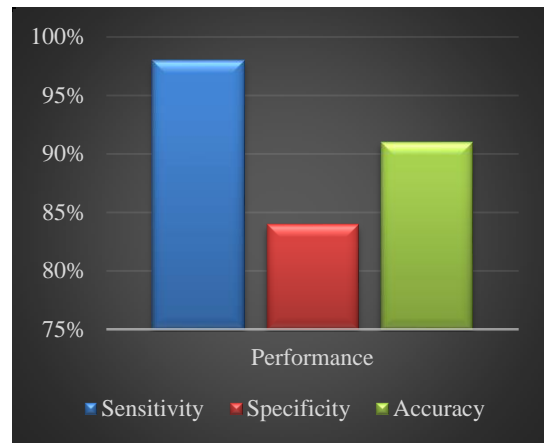


Fig. 20. Classification result.

## REFERENCES

- [1] Sanjay N Rakhade and Frances E Jensen. Epileptogenesis in the immature brain: emerging mechanisms. *Nature Reviews Neurology*, 5(7):380, 2009.
- [2] A.S. Zandi, M. Javidan, G.A. Dumont, R. Tafreshi, Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform, *IEEE Trans. Biomed. Eng.* 57 (7) 1639–1651, 2010.
- [3] L.S. Vidyaratne, K.M. Iftekharuddin, Real-time epileptic seizure detection using EEG, *IEEE Trans. Neural Syst. Rehabil. Eng.* 25 (11), 2146–2156, 2017.
- [4] Wang, L.; Xue, W.; Li, Y.; Luo, M.; Huang, J.; Cui, W.; Huang, C. Automatic Epileptic Seizure Detection in EEG Signals Using Multi-Domain Feature Extraction and Nonlinear Analysis. *Entropy* 2017, 19, 222. <https://doi.org/10.3390/e19060222>.
- [5] Alotaiby, T.N., Alshebeili, S.A., Alshawi, T. et al. EEG seizure detection and prediction algorithms: a survey. *EURASIP J. Adv. Signal Process.* 2014, 183 (2014). <https://doi.org/10.1186/1687-6180-2014-183>.
- [6] H O Lekshmy, Dhanyalaxmi Panickar and Sandhya Harikumar, "Comparative analysis of multiple machine learning algorithms for epileptic seizure prediction", *Journal of Physics: Conference Series*, 2022.
- [7] World Health Organization, *Epilepsy: A Public Health Imperative*. Geneva, Switzerland: WHO, 2019.
- [8] E. Alickovic, J. Kevric, and A. Subasi, "Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction," *Biomedical signal processing and control*, vol. 39, pp. 94–102, 2018.
- [9] Zack M M and Kobau R, National and state estimates of the numbers of adults and children with active epilepsy—United States, 2015 *MMWR Morb. Mortal. Wkly. Rep.* 66 821, 2017.
- [10] Mary Jane England, Catharyn T Liverman, Andrea M Schultz, and Larisa M Strawbridge. *Epilepsy across the spectrum: Promoting health and understanding. A summary of the institute of medicine report.* *Epilepsy & Behavior*, 25(2):266–276, 2012.
- [11] L. E. Hebert, P. A. Scherr, J. L. Bienias, D. A. Bennett, and D. A. Evans, "Alzheimer disease in the US population: prevalence estimates using the 2000 census," *JAMA Neurology*, vol. 60, no. 8, pp. 1119–1122, 2003.
- [12] J. Xiang, C. Li, H. Li, R. Cao, B. Wang, X. Han, J. Chen, the detection of epileptic seizure signals based on fuzzy entropy, *J. Neurosci. Methods* 243, 18–25, 2015.
- [13] D.C. Bergen, do seizures harm the brain? *Epilepsy Curr.* 6 (4), 117–118, 2006.

- [14] Usman SM, Usman M, Fong S. Epileptic Seizures Prediction Using Machine Learning Methods. *Comput Math Methods Med*. 2017.
- [15] Hart, Y. M. Epidemiology, natural history and classification of epilepsy. *Medicine*, 40(9), 471-476, 2012.
- [16] Yousefi, M. R., Golnejad, S., & Hosseini, M. M. (2022). Comparison of EEG based epilepsy diagnosis using neural networks and wavelet transform. *arXiv preprint arXiv:2204.04488*.
- [17] Mattar, E. Application of Machine Learning for Electroencephalography Classification of Epilepsy Seizure Prediction, 2022.
- [18] G. Alarcón and A. Valentín, Introduction to Epilepsy, Cambridge University Press, Cambridge, UK, 2012.
- [19] D. J. Urman, E. Beghi, C. E. Begley et al., "Standards for epidemiologic studies and surveillance of epilepsy," *Epilepsia*, vol. 52, pp. 2–26, 2011.
- [20] R. S. Fisher, "The new classification of seizures by the International League against Epilepsy 2017," *Current Neurology and Neuroscience Reports*, vol. 17, pp. 48–56, 2017.
- [21] Ahmad, I., Wang, X., Zhu, M., Wang, C., Pi, Y., Khan, J. A., ... & Li, G. EEG-based epileptic seizure detection via machine/deep learning approaches: A Systematic Review. *Computational Intelligence and Neuroscience*, 2022.
- [22] Namazi, H., Kulish, V. V., Hussaini, J., Hussaini, J., Delaviz, A., Delaviz, F., & Ramezanpoor, S., A signal processing based analysis and prediction of seizure onset in patients with epilepsy. *Oncotarget*, 7(1), 342, 2016.
- [23] Annuar, R. M. H. R. K., Shahbudin, S., Kassim, M., & Rahman, F. Y. A., Epilepsy Seizure Detection and Classification Analysis using Residual Neural Network. 11th International Conference on System Engineering and Technology (ICSET) (pp. 347-351), 2021.
- [24] Valeta, T. Epileptic Seizures. In *The Epilepsy Book: A Companion for Patients* (pp. 23-32), 2017.
- [25] SS Viglione and GO Walsh. Proceedings: Epileptic seizure prediction. *Electroencephalography and clinical neurophysiology*, 39(4):435–436, 1975.
- [26] Z Rogowski, I Gath, and E Bental. On the prediction of epileptic seizures. *Biological cybernetics*, 42(1):9–15, 1981.
- [27] Mary Ann F Harrison, Ivan Osorio, Mark G Frei, Srividhya Asuri, and Ying-Cheng Lai. Correlation dimension and integral do not predict epileptic seizures. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 15(3):033106, 2005.
- [28] Y Salant, I Gath, and O Henriksen. Prediction of epileptic seizures from two-channel eeg. *Medical and Biological Engineering and Computing*, 36(5):549–556, 1998.
- [29] Wim van Drongelen, Sujatha Nayak, David M Frim, Michael H Kohrman, Vernon L Towle, Hyong C Lee, Arnetta B McGee, Maria S Chico, and Kurt E Hecox. Seizure anticipation in pediatric epilepsy: use of kolmogorov entropy. *Pediatric neurology*, 29(3):207–213, 2003.
- [30] Klaus Lehnertz and Brian Litt. The first international collaborative workshop on seizure prediction: summary and data description. *Clinical neurophysiology*, 116(3):493–505, 2005.
- [31] Florian Mormann, Thomas Kreuz, Ralph G Andrzejak, Peter David, Klaus Lehnertz, and Christian E Elger. Epileptic seizures are preceded by a decrease in synchronization. *Epilepsy research*, 53(3):173–185, 2003.
- [32] Rasheed, K., Qayyum, A., Qadir, J., Sivathamboo, S., Kwan, P., Kuhlmann, L., ... & Razi, A. Machine learning for predicting epileptic seizures using EEG signals: A review. *IEEE Reviews in Biomedical Engineering*, 14, 139-155, 2020.
- [33] Andhale, P., & Patil, D. (2022). Machine Learning Approach for Predicting Epileptic Seizures using EEG Signals: A Review. *Machine Learning Approach for Predicting Epileptic Seizures using EEG Signals: A Review* (February 25, 2022).
- [34] Levin Kuhlmann, Philippa Karoly, Dean R Freestone, Benjamin H Brinkmann, Andriy Temko, Alexandre Barachant, Feng Li, Gilberto Titericz Jr, Brian W Lang, Daniel Lavery, et al. Epilepsyecosystem.org: crowd-sourcing reproducible seizure prediction with long-term human intracranial eeg. *Brain*, 141(9):2619–2630, 2018.
- [35] Matias I Maturana, Christian Meisel, Katrina Dell, Philippa J Karoly, Wendy D'Souza, David B Grayden, Anthony N Burkitt, Premysl Jiruska, Jan Kudlacek, Jaroslav Hlinka, et al. Critical slowing as a biomarker for seizure susceptibility. *bioRxiv*, page 689893, 2019.
- [36] Editor: Bjorn Schelter, Editor: Jens Timmer, and Schulze-Bonhag. *Seizure Prediction in Epilepsy: From Basic Mechanisms to Clinical Application*. John Wiley & Sons, 2008.
- [37] Natu, M., Bachute, M., Gite, S., Kotecha, K., & Vidyarthi, A., Review on epileptic seizure prediction: machine learning and deep learning approaches. *Computational and Mathematical Methods in Medicine*, 2022.
- [38] M. P. Hosseini, T. X. Tran, D. Pompili, K. Elisevich, and H. S. Zadeh, "Multimodal data analysis of epileptic EEG and rs-fMRI via deep learning and edge computing", *Artificial Intelligence in Medicine*, Vol. 104, p. 101813, 2020.
- [39] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, "Focal onset seizure prediction using convolutional networks", *IEEE Transactions on Biomedical Engineering*, Vol. 65, No. 9, pp. 2109-2118, 2017.
- [40] N. M POUR, I.Y ÖZBEK, "Epileptic Seizure Detection based on EEG Signal using Boosting Classifiers" *Erzincan University Journal of Science and Technology*, 14(1), 159-167 2021.
- [41] Prasanna, J., Subathra, M. S. P., Mohammed, M. A., Damaševičius, R., Sairama, N. J., & George, S. T., Automated Epileptic Seizure Detection in Pediatric Subjects of CHB-MIT EEG Database—A Survey. *Journal of Personalized Medicine*, 11(10), 1028, 2021.
- [42] U. Zaki, M. A. Zaki, and S. Narejo, "Chest X-Ray Image-based COVID-19 Recognition using Modified Artificial Neural Network," 3rd International Conference on Computational Sciences and Technologies, MUET Jamshoro, 2022.
- [43] M. A. Zaki, S. Narejo, M. Ahsan, S. Zai, M. R. Anjum, and N. U. Din, "Image-based Onion Disease (Purple Blotch) Detection using Deep Convolutional Neural Network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 5, pp. 448–458, 2021.
- [44] Kulsoom, F., Narejo, S., Mehmood, Z. et al. A review of machine learning-based human activity recognition for diverse applications. *Neural Comput & Applic* 34, 18289–18324 (2022).
- [45] S. Imtiaz, S. -F. Horchidan, Z. Abbas, M. Arsalan, H. N. Chaudhry and V. Vlassov, "Privacy Preserving Time-Series Forecasting of User Health Data Streams," 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 2020, pp. 3428-3437, doi: 10.1109/BigData50022.2020.9378186.
- [46] Narejo, Sanam; Pasero, Eros; Kulsoom, Farzana (2016). EEG Based Eye State Classification using Deep Belief Network and Stacked AutoEncoder. In: *International Journal of Electrical and Computer Engineering*, vol. 6 n. 6, pp. 3131-3141. - ISSN 2088-8708.

# The Impact of COVID-19 on Digital Competence

## A Case Study of Preservice Teacher Education Students in the Sultanate of Oman

Syerina Syahrin<sup>1</sup>, Khalid Almashiki<sup>2</sup>, Eman Alzaanin<sup>3</sup>

College of Arts and Applied Sciences, Dhofar University, Sultanate of Oman<sup>1,2</sup>

Faculty of Languages and Translation, King Khalid University, Saudi Arabia<sup>3</sup>

**Abstract**— The study looked into how COVID-19 affected the digital competence of a group of preservice teacher education students at a higher education institution in the Sultanate of Oman. The paper examined students' digital profile in five areas namely information and data literacy, communication and collaboration, digital content creation, safety and problem solving. Data from 32 undergraduate students was collected by utilizing DigComp, a European Commission digital skills self-assessment tool and findings from a survey. The digital competence framework measures the set of skills, knowledge and attitudes that describes what it means to be digitally competent. These skills are important for students to be effective global citizens in the 21<sup>st</sup> century. The results of the study revealed that the majority of the students scored Level 3 (Intermediate) in their self-assessment competency test score. The majority of the students perceived that their digital competence improved significantly as the result of online learning which was accelerated by the COVID-19 pandemic. The rationale of this investigation is that it helps educators understand the students' level of digital competence and the students' perspectives on ICT skills. In turn, it informs us the ways to monitor the students' digital progress and the next steps in developing their digital competency.

**Keywords**—Digital competence; digital skills; digital profile; ICT skills; preservice teacher education

### I. INTRODUCTION

Oman's National Education Strategy 2040 aims to align the country's education system with the changing needs of the economy and society, with a focus on developing the skills and knowledge that student need to be successful in a rapidly changing world. This includes a focus on critical thinking, digital literacy, as well as the promotion of values such as creativity, innovation, and entrepreneurship [58]. The Education Council explained that twenty-first century skills are the "global currency" (p.17) for local and international workplaces [52]. Developing digital competencies can help students in Oman to develop skills and competencies that are in line with the needs of international labor markets. Digital competencies such as proficiency in technology, digital literacy, and digital communication are becoming increasingly important in today's global economy. By developing these skills, students in Oman can better prepare themselves for jobs in various fields, including technology, business, and communication. Additionally, digital competencies can also help students to become more adaptable and resilient in an ever-changing job market [2]. Many jobs today need digital skills [28]. To that end, the Education Council [52] described:

Oman's movement towards a knowledge-based economy requires reform of the current education system. [It] should consider improving the information technology infrastructure in educational institutions [and] supporting the use of cutting-edge technology, e-learning, capacity building and infrastructure improvement in schools and higher education (p.33).

The COVID-19 pandemic led to a rapid shift towards digital learning as universities were forced to close and shift to remote learning. This included the use of online learning platforms, video conferencing tools, and virtual classrooms. In a recent report, the Ministry of Transport, Communications, and Information Technology has set-up plans and strategies to equip 10,000 young Omanis with digital skills by 2025 [4]. The task of learning new digital skills is not new in Oman. Prior to the pandemic, students in the higher education institutions in Oman has somewhat experienced using electronic databases and e-learning platforms, as well as engaging with their peers in formal learning context online [38]. However, students often cited internet speed and internet coverage as factors that limit their online experience [50]. However, the internet speed, bandwidth and coverage in the Sultanate of Oman dramatically increased after the pandemic [41]. As a result, universities, such in the case of this study, Dhofar University, were able to adapt to the challenges posed by the pandemic and continue providing education to students through online learning [49]. As the result of that, the instructors at the university were able to utilize synchronous and asynchronous classroom instructions during online learning. Dhofar University is a private higher education institution located in the southern region of Oman. The Methodology section of a study provides a detailed explanation of the research design and methods used to conduct the study.

The COVID-19 crisis has led to the discovery of new and creative ways to utilize technology for teaching, learning, and analysis. The research aims to investigate the effect of post-pandemic online learning on the digital competence of a group of students in preservice teacher education. The study is framed by the following research questions:

1) How do the preservice teacher students evaluate their digital competencies?

2) How do the preservice teacher students perceive their competencies in Information and Communications Technology (ICT) post-pandemic?

3) What is the relationship between the number of years the preservice teacher students report to actively use ICT in their everyday lives to their digital assessment score?

The present study is significant as it documents the preservice teacher students' digital profile and sought the students' perspectives on their ICT skills. The individuals examined in the study were students preparing to become teachers. Currently, there is a lack of knowledge about the digital abilities of these student teachers within the context of Oman. The World Bank reported two important findings that set the backdrop of this study. First, the majority of Omani teachers have less than a decade of teaching experience. Second, the majority of the teachers do not know how to incorporate technology in the classroom. The lack of digital competence among the new teachers is a cause for concern. As Oman shifts its focus to hiring local talent through its Omanization policy, it is crucial to ensure that preservice teachers, like the participants in this study, are proficient in digital skills. Omanization is the national initiative to increase the employment of local citizens in the workforce [2]. Oman Vision 2040 has placed the education sector at the center of the country's development. In order to ensure that the country produces highly competent teachers, it is evident that teacher education programs need to be scrutinized.

The study aimed to gather information on the students' digital profiles and their views on their ICT skills before and after the online learning during the COVID-19 pandemic, using a survey method. The main idea behind the research was that online learning would result in an improvement in self-perceived digital proficiency.

## II. BACKGROUND OF STUDY

### A. Digital Competence in Higher Education

The widespread use of information and communication technology (ICT) has made digital competence a commonly discussed topic in academic literature and policy documents [60] [7]. The rise of digitalization has also led to new expectations in higher education [60]. The heavy influence of digitalization can be seen in the way it has impacted the knowledge development process and the performance of both teachers and students.

Digital competence is a clear priority in European policies [47], hence the common references to European Union policy documents when defining the concept in academic publications [60]. In the context of higher education, digital competence is poorly defined, and most scholars use definitions from policy papers [60]. A viable explanation is the future-oriented nature of digital competence, as it focuses on the skills required to become a professional and citizen [47]. Higher educational institutions are responsible for the development of these skills among students [47]. This view is echoed in the United Nation's 2030 Agenda for Sustainable Development [57], which endorsed that ICT skills and competencies must be harnessed by young people, and that digital technologies enable rethinking teaching and learning processes. There is a growing place for ICT and digital competence in and out of the HE classroom to support teaching and learning [44].

### B. Teachers' Digital Competencies in Higher Education

Digital competence is among the key challenges that teachers in higher education face [8]. However, there is no established definition for professional digital competence, nor is it an established practice [9]. Although digital competence is becoming essential for citizens in general, teachers in higher education specifically hold a responsibility to master it. Digital transformation and improvement of education very much depend on teachers' actions [8]. Professional digital competence in education settings is highly contextual, as teachers need to manage complexity and master decision-making for the optimal use of ICT to meet learning outcomes [9] [53]. Moreover, studies recognize that teachers are role models for students, who need to be digitally confident to inspire the class to make the best use of ICT [5] [26].

A popular conceptualization of teachers' digital competence is the layered approach levels [26]. The macro level is concerned with generic skills. At the meso level is didactic digital competence and how classrooms are affected and afforded by ICT. The micro level is connected to teachers' enactment of digital competence to design courses, communicate with audiences and stakeholders manage digitally-rich classrooms [22]. A fourth dimension, suggesting that educators should possess transformative agency [9]. The term captures the notion that teachers are not passive recipients of frameworks and tools. Instead, they take the initiative to transform practices and engage in problem-solving by implementing ICT, depending on the classroom situations they face [9] [1].

### C. Impact of the COVID-19 Pandemic on Preservice Teachers' Digital Competence

It was against this backdrop when the COVID-19 pandemic abruptly replaced most face-to-face teacher education courses with online lectures [42] [45]. The pandemic pushed the importance of digital competence in higher education to a new height [60]. Pre-service teachers had to go through their training online [25]. It is not that they were not familiar with technology like Google Classroom, Zoom or WhatsApp. However, some of them were not ready for extensively using it [6]. Moreover, student teachers also had to deliver online lectures as a part of their practicum [6] [10].

On the one hand, the coronavirus outbreak encouraged teachers in higher education to quickly change their practices to allow for learning continuity from distance [8]. On the other hand, teachers' urgent response to the crisis involved little time for preparation [25] [35]. A few studies posed the question of whether the digital competence of pre-service teachers increased during the pandemic or not [24] [29]. A study found that during the first year of the pandemic prospective teachers at a German university made enormous changes towards digital technology adoption [24]. Despite this, the researchers did not report a significant impact on their digital competencies, probably because they were seeing the pandemic as a temporary event and were planning to go back to their pre-pandemic behaviour. Another study surveyed 147 pre-service teachers to reveal their self-reported digital literacy pre-university, or before the pandemic, and during

distance education after the COVID-19 outbreak [43]. Findings show that pre-university evaluations were more positive, while the emergency highlighted some weaknesses that resulted in lowered evaluations of digital literacy. Much of the issue stems from widespread the use of smartphones to attend online classes, which causes distractions. Nevertheless, 27% of participants reported improvement in digital communication skills [43]. Another report looked at the written reflections of 51 student teachers in Norway [25]. Their findings demonstrate an overall reluctance to integrate digital technology during teaching practicum due to some difficulties related to digital competence. The scholars highlight the need for continuous development of ICT skills, as the pandemic has demonstrated their critical significance for teachers' education.

The Background section addresses issues that support the need for a research design that allows preservice teacher education students to evaluate their digital competence and ICT skills before and after the pandemic.

### III. METHODOLOGY

#### A. Context of the Study

The Sultanate of Oman is situated in the southeast portion of the Arabian Peninsula and shares borders with Saudi Arabia, the United Arab Emirates, and the Republic of Yemen. The latest census shows that the population of Oman is slightly above 4.5 million, with Omani citizens making up 62% of the population and expatriates comprising 38% [37]. The majority of the study participants come from the Dhofar region, which is located in the southern part of the country and borders the Republic of Yemen.

The Sultanate of Oman generates most of its revenue from oil resources and it has been the source of the country's development [46]. After oil, the fisheries account for the majority of the country's trade export. The government identified other sectors as means to diversify the country's economy. The sectors are tourism, fisheries, mining, logistics and manufacturing. The Sultanate also aimed to make the ports in the country into regional maritime hubs [46]. The traditional economic activities in Oman, in particular the Dhofar region, are production of honey, frankincense, fisheries, and agriculture which includes raising animal livestock such as camels and cows [37]. While a proportion of the population still maintain the traditional economic activities, some moved and worked in the cities to take up jobs in various government and private sectors [37]. The government is also pushing the agenda for the growth of entrepreneurship and Small-Medium Enterprises (SME) for wider socio-economic goals, as part of the country's Vision 2040 [27]. The national policy also outlined digitization and technological innovation in transforming the country's economy [37].

In the education sector, the ministry is rolling-up the plans for digital transformation. This included training teachers with skills and technology [38], developing IT infrastructure [40], and equipping students with digital skills for their future careers [4]. Being digitally competent in all areas of work, in

particular, the education sector is crucial in the development of the Sultanate of Oman.

#### B. Participants of the Study

The study recruited 32 undergraduate students (28 female, 4 male) from Dhofar University in Oman, using a systematic sampling method. These participants were all enrolled in a preservice teacher course and had prior experience with online learning due to the pandemic. The participants were at the second, third, or fourth year level of their undergraduate studies and had the intention to become teachers after graduation.

#### C. Data Collection

The participants of the study are 32 undergraduate preservice teacher education students at Dhofar University. Dhofar University students who participated in the research experienced remote learning during the restricted movement as part of the country's initiatives to minimize the spread of COVID-19 pandemic from March 2020 to July 2021. Although some of the students had experienced some aspects of technology in the classroom prior to the pandemic, the abrupt move towards online learning forced the students to acquire new skills without sufficient training and in a ramp-up time [59].

Dhofar University's learning management system is Moodle. Through Moodle, the students received synchronous and asynchronous instructions. The instructors at Dhofar University, at least some of them, used video conferencing tools to deliver their classes virtually. There were also several instructors who used game-based learning tools. The students who participated in the research returned to campus for face-to-face learning beginning September 2021. Although the majority of the courses are delivered face-to-face, some instructors continued to incorporate some aspects of technology in their classroom teaching and learning.

The data was gathered and analyzed using the following steps. First, the students were invited to evaluate their digital competence in the form of an online test. The testing tool that was used, as further explained in the Research Instruments section, is DigComp. The data gathered from the test were recorded as their digital profiles in five areas namely, information and data literacy, communication and collaboration, digital content creation, safety and problem solving. All items were averaged respectively. The data was then interpreted using Dimension 3 Proficiency Level from the European Digital Competence Framework for Citizens. Second, the students were invited to self-assess their ICT competencies pre- and post- pandemic through an online survey. The survey items were adapted from the DiCTE project [15]. The hypothesis being tested was that participating in online learning during the pandemic would result in a rise in self-reported digital proficiency. The statistical analysis was carried out using SPSS (Version 27). To evaluate the hypothesis, a Paired-Samples test was applied to the mean of responses to all items ( $p < .05$  significance level). Additionally, the open-ended responses were analyzed, coded and grouped according to themes.

#### D. Research Instruments

The investigation utilizes two research instruments. The first is the testing instrument called the DigCom. The instrument is developed from the European Digital Competence Framework for Citizens DigComp 2.1. [19]. The digital competence framework measures the set of skills, knowledge and attitudes that describes what it means to be digitally competent. These skills are important for students to be effective global citizens in the 21st century. The test covers five competence areas namely information and data literacy, communication and collaboration, digital content creation, safety and problem solving [20]. When completing the test, the participants will receive overall results per competence area and an average level for all competences (Foundation, Intermediate or Advanced). The test allows participants to evaluate their own digital competence level, from 1 to 6. Levels 7 and 8 of DigComp 2.1 are not taken into consideration as they are reserved for highly specialized careers [20]. DigCom has been used and tested in many countries in Europe as part of their Digital Education Plan, as well as countries outside of Europe and this can be traced in the report by the Council of European Professional Informatics Societies (CEPIS) [11] [51]. In this study, the DigComp test results are used as a reference to define and assess digital profiles for the preservice teacher education students.

The online survey serves as the second research tool. It enables students to assess their proficiency in ICT, from “not at all” to “very high degree”, both pre- and post-pandemic. The survey was adapted from the DiCTE project, which stands for Developing ICT in Teacher Education and is funded by the European Commission. The survey created by the DiCTE project aims to determine the digital competence of student teachers at the start of their education [15] [31].

#### IV. RESULTS

DigCom test covers five competence areas namely information and data literacy, communication and collaboration, digital content creation, safety and problem solving, as indicated in Table I below. The data are then interpreted using Dimension 3 of the European Digital Competence Framework.

TABLE I. DIGCOM MEAN SCORE

Competencies	Mean score
Information and data literacy	2.451612903
Communication and collaboration	2.709677419
Digital content creation	3.387096774
Safety	2.290323
Problem-solving	2.193548387
Overall	2.606452

The results from Table I revealed the mean score of the participants’ digital profile is as follows:

- 1) Information and data literacy at the mean score of 2.45

- 2) Communication and collaboration at the mean score of 2.7
- 3) Digital content creation at the mean score of 3.38
- 4) Safety at the mean score of 2.29
- 5) Problem-solving at the mean score of 2.19

The overall mean score of the students’ digital competence in all five areas is at 2.6. Each of the mean score is interpreted using Dimension 3 of the European Digital Competence Framework. Each of the competency is explained in turns.

##### A. Information and Data Literacy

The students’ competency in Information and Data Literacy is based on their ability in the following categories:

- 1) Browsing, searching, and filtering data
- 2) Evaluating data
- 3) Managing data, information, and digital content

The preservice teacher education students who participated in this study scored an average of 2.45 for Information and Data Literacy. According to the Dimension 3 of the European Digital Competence Framework, the students’ competence in Information and Data Literacy is in the Foundation 2 category. This means, the students are able to do a simple search on their own or with guidance. They can identify their information needs, detect the credibility and reliability of data sources, and organize information in structured environments.

##### B. Communication and Collaboration

The participants’ competence in Communication and Collaboration is at the mean score of 2.7. The students’ competency in Communication and Collaboration is based on their ability in the following categories:

- 1) Interacting through digital technologies
- 2) Sharing through digital technologies
- 3) Engaging citizenship
- 4) Collaborating through digital technologies
- 5) Netiquette
- 6) Managing digital identity

The students’ proficiency in Communication and Collaboration is categorized at the level of Intermediate 3. Dimension 3 of the DigComp 2.1 framework described the students’ competence at Intermediate 3 Communication and Collaboration as the ability to perform interactions with digital technologies, select appropriate digital communication to participate in society, to empower oneself and to collaborate with others. At this level too, students are able to explain know-how while using technologies, express communication strategies adapted to an audience, and ways to protect one’s reputation online.

##### C. Digital Content Creation

The participants’ digital competence in Digital Content Creation is at the mean score of 3.38. Out of the five digital competence areas, this area is the highest mean score. When interpreted against Dimension 3 of the European Digital Competence Framework, the score can be categorized as Intermediate 3 level.

The participants' digital competence in Digital Content Creation Literacy is based on their ability in the following categories:

- 1) Developing digital content
- 2) Integrating and re-elaborating digital content
- 3) Copyright and licenses
- 4) Programming

According to Dimension 3 of the DigComp 2.1 framework, at Intermediate 3 Digital Content Creation, the students are able to create and edit routine content, express oneself through the digital content creation, indicate rules of copyright and licenses, as well as to list instructions for a computing system to solve routine problems.

#### D. Safety

The participants' digital competence in Safety is at the mean score of 2.29. When interpreted against Dimension 3 of the European Digital Competence Framework, the score can be categorized at the Foundation 2 level. The students' digital competency in Safety is based on their ability in the following categories:

- 1) Protecting devices
- 2) Protecting personal data and privacy
- 3) Protecting health and well-being
- 4) Protecting the environment

At the Foundation 2 level, students are able to identify simple ways to protect their devices and digital content, differentiate simple risks and threats, follow simple safety and security measures, and have due regard to reliability and privacy. They can also select simple ways to protect their personal data and privacy in digital environments, use and share personally identifiable information while protecting themselves and others from damages, differentiate simple ways to avoid health risks and threats to physical and psychological well-being while using digital technologies, and select simple ways to protect themselves from possible dangers in digital environments [20].

#### E. Problem-Solving

Lastly, the participants' digital competence in Problem-Solving is at the mean score of 2.19. Out of the five digital competence areas, this is the lowest mean score. When interpreted against Dimension 3 of the European Digital Competence Framework, the score can be categorized at the Foundation 2 level. The students' digital competency in Problem-Solving is based on their ability in the following categories:

- 1) Solving technical problems
- 2) Identifying needs and technological responses
- 3) Creatively using digital technology
- 4) Identifying digital competence gaps

At the Foundation 2 level, students can recognize basic technical issues while using devices, propose basic solutions, identify needs, and understand basic digital tools and how they can be used to address those needs. They can also identify digital tools and technologies that can be used to generate new knowledge and improve processes and products. Additionally, they are able to address basic problems in digital

environments, recognize areas where their own digital skills need improvement or updating, and identify opportunities for self-development and staying current with digital advancements [20].

#### F. Overall Digital Competence Score

The overall mean score of the participants' digital competence is at 2.6. When interpreted against the Dimension 3 of the European Digital Competence Framework, we can round-up the mean score and conclude that the students' digital competence can be categorized as Intermediate 3 level.

#### G. ICT Competencies Pre- and Post- Online Learning COVID-19 Pandemic

The results for the next section are presented based on the statistical tests outlined in the Data Collection section. Significance levels are indicated by \* for  $p < .05$ , \*\* for  $p < .01$ , and \*\*\* for  $p < .001$ .

The students self-assessed their ICT competencies pre- and post- pandemic in regard to the following areas:

- 1) Competence in using word processor
- 2) Competence in using spreadsheet
- 3) Competence in using Presentation tools
- 4) Competence in using photo and video editing applications and tools
- 5) Competence in using Learning Management System
- 6) Competence in using educational games

The research utilized paired samples test to measure the differences in the way the students perceived the impact of online learning on their digital competencies, pre- and post-pandemic. The following hypotheses guided the statistical analysis of the study:

The null hypothesis ( $H_0$ ):  $\mu =$  there is no association of the students' participation in online learning to an increase in self-assessed digital competence

The alternative hypothesis: ( $H_a$ ):  $\mu =$  there is an association of the students' participation in online learning to an increase in self-assessed digital competence

The data suggests that the students perceived an increase in ICT competence post- pandemic, as they participated in online learning. Utilizing a 0.05 level of significance, the p-values are of 0.001, 0.000 and 0.002 are less than the significance level of 0.05. Thus, it rejects the null hypothesis.

The common themes emerged from the open-ended responses to, "as a preservice teacher education student, I should be digitally competent because ..." are coded and categorized as follows:

- 1) Learning approach for present and future times
- 2) Prepare students with skills for the workplace
- 3) Prepare students for the global society

Finally, the findings measured the correlation between the participants' digital competency overall score (DigCom) to the number of years they reported to actively use ICT in their everyday lives. The Pearson correlation coefficient for the number of years and score is .893, which is significant ( $p < .01$  for a two-tailed test). There is a significant relationship



between years of experience and competency score. It demonstrates that the student's experience in utilizing technology in their everyday lives has a significant impact on their competency score, as illustrated in Table II below.

TABLE II. CORRELATION BETWEEN NUMBER OF YEARS ACTIVELY USE ICT AND DIGITAL COMPETENCY SCORE

		Score	Number of years
Score	Pearson Correlation	1	.893**
	Sig. (2-tailed)		.000
	N	32	32
Number of years	Pearson Correlation	.893**	1
	Sig. (2-tailed)	.000	
	N	32	32

\*\* . Correlation is significant at the 0.01 level (2-tailed).

## V. DISCUSSION

The paper investigated the perspectives of a group of preservice teacher education students on their digital competence by assessing their digital competence before and after the online learning resulted from the pandemic. In regards to the first research question, the findings show the students' overall proficiency in digital skills is at Intermediate 3. This finding is similar with the result with undergraduate Chilean students, with a sample size of 817 participants [48]. Although the overall score is similar, the Chilean report differs when it comes to the five digital areas. The Chilean students scored highest for Safety/Network Security and lowest for Digital Content Creation. The students participated in this research on the other hand, scored highest for Digital Content Creation. In both countries, Chile and the Sultanate of Oman, the low mean score for Problem-Solving is a cause for concern. In current scholarship, particularly in the context of the Sultanate of Oman, the inability of graduates in solving problems is viewed as one of the major educational challenges [36]. The educational policy makers and curriculum developers in the country were tasked to ensure that soft skills, such as problem-solving are integrated in the curriculum. Previous investigations addressed different ways problem-solving skills may be incorporated in the classroom teaching and learning [55] [3].

An area that needs to be addressed is the participants' increased competency in Digital Content Creation, particularly when it comes to developing and integrating digital content. One possible explanation is that the participants are active and frequent users of Web 2.0. Web 2.0 refers to the current state of the internet technology that allows users to generate content for other end users [13]. The earlier version of the internet, Web 1.0 had fewer content creators and the majority of the internet users were only consumers [13]. Web 2.0 gave rise to various social media platforms. The skills associated with the active and frequent use of social media may contribute to the reported level of skills in creating digital contents. With the population of slightly above 4.5 million [39], social media applications in the country are Facebook and Instagram. Facebook has 1.6 million users while Instagram has 1.2 million users [39]. The popularity of the social media

applications depends on the active participation of the users in producing, sharing, and interacting about digital contents.

In regards to the second line of inquiry, the research measured the differences in the perceived ICT competence pre- and post- pandemic. The finding rejected the null hypothesis and unveiled that there was an association of the students' participation in online learning to an increase in self-assessed digital competence. In their self-assessment, the participants reported to have increased ICT competence in using word processor, spreadsheet, presentation tools, photo and video editing tools, Learning Management System and educational games post- pandemic, resulted from online learning. Similar finding was traced in previous studies [16] [18]. The study found that the preservice teachers' competence and online teaching readiness was enhanced post-pandemic. Although the findings favored online learning as it has positively impacted the students' ICT skills, we cannot assume that the students are prepared with pedagogical practices to teach online. In a study in the US, the authors found that even though most teachers perceived readiness, they did not use the appropriate digital tools when teaching in the classroom [16].

Another consideration is regarding the survey instrument. The survey' results represent the students' belief about their ICT competence rather than their actual level of competence. This latter should be measured by performance tasks [21]. Previous research explained that one of the weaknesses of survey instruments is the respondents themselves [30]. The participants may respond to the survey items with social-desirability bias such as the desire to please the instructor or bias in self-perception. The previous study addressed the need for the validation of the survey instruments, particularly when it is used as the yardstick in measuring one's competence [54].

The open-ended item in the survey sought the participants' perspectives on the importance of equipping oneself with digital skills. The common themes that emerged from the responses can be summarized and categorized as follows:

- The participants linked equipping oneself with digital skills to the ability to approach classroom teaching and learning that is relevant for present day and for the future
- The participants linked the importance of equipping oneself with digital skills to prepare students for the workplace and global society.

An excerpt from one of the responses states,

"It's important for me to be digitally competent so that I can impart my knowledge to my students in a new way. We are no longer living in a traditional way- to learn from books. Students today learn from YouTube and Google". (S11)

The excerpt above is one of the examples the participants gave in which the student linked the importance of preservice teacher education students equipping themselves with ICT skills to the ability to approach the classroom that fit current and future times. In other words, digitally competent teachers are able to make classroom teaching and learning relevant.

Other excerpts state,

“I don’t see how we can teach without having some ideas about technology. Maybe we are not IT experts, but we should have some ideas. In future, our students need IT skills for their work too. It starts with us”. (S8)

“Oman is no longer a traditional society. We interact with people from different countries through studies, work, government and so on. Being good with technology is important for us as future teachers so that we can show the younger Omani children how to use technology to connect with people from other countries”. (S3)

The excerpts above are examples the participants gave in which the students linked the importance of preservice teacher education students equipping themselves with ICT skills to preparing students for the workplace and the global society. Being competent in ICT is an unarguably an essential part of teacher education [23]. In a study with preservice teacher education in Ireland, Norway and Spain, the authors found that despite the different ways the education policies developed, the participants had a similar and shared view of the importance of digital skills [32]. Another study found that preservice teacher education students in an Irish University had a favorable positive attitude toward using technology in the classroom [33]. Both of the studies utilized the DiCTE project survey as their instrument of study.

Finally, this paper investigated the relationship between the participants’ digital competency score, which was derived from DigComp digital profile to the participants’ reported active use of ICT in their daily lives. The active use of ICT is narrowed into the following: the use of word processor, spreadsheet, presentation tools, photo and video editing tools, Learning Management System and educational games. The results found there is a strong correlation between the number of years the participants reported to actively use ICT in their lives to their digital competence score. In other words, the duration of exposure and active use of ICT has a significant impact on their digital competencies. This discovery aligns with previous research that found teachers who feel confident in their ability to use educational technology and see it as beneficial for teaching and learning are more likely to incorporate digital tools in their teaching. Similarly, providing preservice student teachers with more opportunities to work with digital tools has been found to improve their ability to use technology in their classrooms in the future [12] [14].

## VI. CONCLUSION

The study examined the effects of COVID-19 on the digital skills of preservice teacher education students at Dhofar University in Oman, focusing on the students’ level of digital competence, their views on their ICT abilities post-pandemic, and the correlation between their reported years of active ICT use and their digital competency scores.

The study utilized two instruments that were DigCom, a digital skills self-assessment tool which was produced by the European Commission and a survey which was adapted from the DiCTE project. Both of the research instruments have been validated by studies around the world, as explained in the Discussion section.

In summary, the findings first revealed that the majority of the students scored Level 3 (Intermediate) in their self-assessment competency test score. The results of the study also showed that among the five areas tested, namely information and data literacy, communication and collaboration, digital content creation, safety and problem solving, the students scored the highest for digital content creation and the lowest for problem-solving. Second, the findings showed that the majority of the students perceived that their digital competence improved significantly as the result of online learning which was accelerated by the COVID-19 pandemic. Finally, the study demonstrated that the duration of active use of ICT has a significant impact on the students’ digital competency score.

The rationale of this investigation is that it helps educators understand the preservice teacher students’ level of digital competence. It also gave us an insight on the ways the students perceived their ICT skills before and after the pandemic. In turn, the data informs us the ways we could identify their strengths and weaknesses, monitor their progress and identify ways in supporting and developing their digital skills.

An avenue for future research, is to consider measuring the students’ digital competency based on performance tasks. The present study is based on the students’ self-assessment, which reflect their belief about their digital competence. Another consideration to make is to have digital assessment that is more relevant to the Arab region. The expression of survey result may differ, as the participants respond to statements that are more contextually relevant to them.

The implication that can be drawn from the findings of the study is that, if the students were given more time and opportunities to engage with various forms of ICT, the more they would perceive themselves to be digitally competent. This belief would then influence the way they conduct their classes. The more positive the belief is, the more effectively they would incorporate aspects of technology in their classroom. As future teachers, the students have a greater responsibility to empower their students with important skills for work and participation in the society through the use of technology. As the DiCTE project [15] explains,

“Formal education has an important role to play in this regard and none more so than in teacher education. From a perspective of equity, it is crucial for teacher education to equip future teachers with the required professional digital competence so that they can prepare children to grow up in a digital society” (para.2).

The growing importance of technology in education is without a doubt, transforming the ways learning takes place. In the Sultanate of Oman, there have been suggestions for hybrid education in which combine the aspects of face-to-face and online learning in classroom teaching and learning [34]. Reimagine Education, an initiative pioneered by the UNICEF explained that technology could bring about change in delivering learning opportunities to children and young people “anywhere, at any time” [56]. The growth of technology requires innovative pedagogies, which could only take place if the future teachers are digitally competent. As these students

would be future teachers, they would be given a greater responsibility in designing instructional strategies that align with the country's Vision 2040, the implications of this study are immediate to them.

#### REFERENCES

- [1] Agyei, D. D., & Voogt, J. M. (2011). Exploring the potential of the will, skill, and tool model in Ghana: Predicting prospective and practicing teachers' use of technology. *Computers & Education*, 56(1), 91–100. <http://dx.doi.org/10.1016/j.compedu.2010.08.017>
- [2] Al Hasani, S. H., & Husin N. A. (2021). A review of digital transformation of education in Oman. *Journal of Business Management and Accounting*, 11(2), 41-59. <https://doi.org/10.32890/jbma2021.11.2.3>
- [3] Al-Mahrooqi, R., & Denman, C. J. (2020). Assessing students' critical thinking skills in the humanities and sciences colleges of a Middle Eastern university. *International Journal of Instruction*, 13(1), 783-796. <http://dx.doi.org/10.29333/iji.2020.13150a>
- [4] Al-Mashaani (2022, September 20). Plan to equip 10,000 young Omanis with digital skills by 2025. *Oman Daily Observer*. <https://www.omanobserver.om/article/1125510/business/plan-to-equip-10000-young-omanis-with-digital-skills-by-2025>
- [5] Alnasib, B.N.M. (2023). Digital competencies: Are pre-service teachers qualified for digital education?. *International Journal of Education in Mathematics, Science, and Technology*, 11(1), 96-114.
- [6] Ayumi, J. S., Rezeki, Y. S., & Wardah, W. (2022). EFL pre-service teachers' experiences doing practicum during COVID-19 pandemic. *The Journal of English Literacy Education: The Teaching and Learning of English as a Foreign Language*, 9(1), 65-77.
- [7] Bartolomé, J., de Soria, I. M., Jakobsone, M., Fernández, A., Ruseva, G., Koutoudis, P., ... & Vaquero, M. (2018, March). Developing a digital competence assessment and accreditation platform for digital profiles. In Proceedings of the 12th International Technology, Education and Development Conference (INTED), Valencia, Spain (pp. 5-7).
- [8] Basilotta-Gómez-Pablos, V., Matarranz, M., Casado-Aranda, L. A., & Otto, A. (2022). Teachers' digital competencies in higher education: a systematic literature review. *International Journal of Educational Technology in Higher Education*, 19(1), 1-16. <https://doi.org/10.1186/s41239-021-00312-8>
- [9] Brevik, L. M., Gudmundsdóttir, G. B., Lund, A., & Strømme Aanesland, T. (2019). Transformative agency in teacher education: Fostering professional digital competence. *Teaching and Teacher Education: An International Journal of Research and Studies*, 86. <http://dx.doi.org/10.1016/j.tate.2019.07.005>
- [10] Çebi, A., & Reisoğlu, İ. (2020). Digital competence: A study from the perspective of pre-service teachers in Turkey. *Journal of New Approaches in Educational Research*, 9(2), 294-308.
- [11] CEPIS (n.d.) *Awareness of the European digital competence framework for citizens (DigComp) framework among the CEPIS community*. The Council of European Professional Informatics Societies (CEPIS). <https://cepis.org/digcomp-report-2021/>
- [12] Chiara A., Alberto C., & Francesca A. (2022) Can teachers' digital competence influence technology acceptance in vocational education? *Computers in Human Behavior*, 132. <https://doi.org/10.1016/j.chb.2022.107266>.
- [13] Conole, G., & Alevizou, P. (2010). A literature review of the use of Web 2.0 tools in higher education. Higher Education Academy.
- [14] Costley, K. C. (2014). The Positive Effects of Technology on Teaching and Student Learning. *Arkansas Technology University*.
- [15] Developing ICT in Teacher Education (DiCTE) (n.d.). *About the DiCTE project*. <https://dicte.oslomet.no/about-the-dicte-project/>
- [16] Ersin, P., Atay, D., & Mede, E. (2020). Boosting preservice teachers' competence and online teaching readiness through e-practicum during the COVID-19 outbreak. *International Journal of TESOL Studies*, 2(2), 112-124. <http://dx.doi.org/10.46451/ijts.2020.09.09>
- [17] European Commission (2018) DigComp: The European Digital Competence Framework.
- [18] European Commission (2020). *Resetting education and training for the digital age*. Luxembourg: Publications Office of the European Union
- [19] European Commission (n.d.) *Test your digital skills!* <https://europa.eu/europass/digitalskills/screen/home?referrer=epass&route=%2Fen>
- [20] European Commission, Joint Research Centre, Vuorikari, R., Kluzer, S., Punie, Y. (2022). *DigComp 2.2, The Digital Competence framework for citizens: with new examples of knowledge, skills and attitudes*, Publications Office of the European Union. <https://data.europa.eu/doi/10.2760/115376>
- [21] Guàrdia, L., Crisp, G., & Alsina, I. (2017). Trends and challenges of e-assessment to enhance student learning in Higher Education. *Innovative practices for higher education assessment and measurement*, 36-56. <http://dx.doi.org/10.4018/978-1-5225-0531-0.CH003>
- [22] Gudmundsdóttir, G. B., & Hatlevik, O. E. (2018). Newly qualified teachers' professional digital competence: implications for teacher education. *European Journal of Teacher Education*, 41(2), 214e231. <https://doi.org/10.1080/02619768.2017.1416085>
- [23] Gudmundsdóttir, G. B., Gassó, H. H., Rubio, J. C. C., & Hatlevik, O. E. (2020). Student teachers' responsible use of ICT: Examining two samples in Spain and Norway. *Computers & Education*, 152, 103877. <https://doi.org/10.1016/j.compedu.2020.103877>
- [24] Höfer-Lück, H., Delere, M., & Vogel, T. (2020). Changing Practices and Self-Reflection? Implications of the Corona Crisis Regarding Private and Professional Digital Media Use of Pre-Service Teachers. In EDULEARN20 Proceedings. 12th International Conference on Education and New Learning Technologies (pp. 7319-7327).
- [25] Jimarkon, P., Wanphet, P., & Dikilitas, K. (2021). Pre-service Teachers' Digital Experiences through Digital Pedagogical Practices in Norway. *Nordic Journal of Comparative and International Education*, 5(4), 86-103.
- [26] Krumsvik, R. J. (2008). Situated learning and teachers' digital competence. *Education and information technologies*, 13(4), 279-290. <https://doi.org/10.1007/s10639-008-9069-5>
- [27] Kutty, S. (2022, June 6). Small but mighty. *Oman Daily Observer*. <https://www.omanobserver.om/article/1120442/business/markets/small-yet-mighty>
- [28] Leahy, D., & Wilson, D. (2014, July). Digital skills for employment. In *IFIP Conference on Information Technology in Educational Management* (pp. 178-189). Springer, Berlin, Heidelberg. [http://dx.doi.org/10.1007/978-3-662-45770-2\\_16f](http://dx.doi.org/10.1007/978-3-662-45770-2_16f)
- [29] Marchisio, M., Barana, A., Fissore, C., & Pulvirenti, M. (2021). Digital Education to Foster the Success of Students in Difficulty in Line with the Digital Education Action Plan. In *Lessons from a pandemic for the future of education-EDEN Conference 2021* (No. 1, pp. 353-363). European Distance and E-learning Network. <https://doi.org/10.38069/edenconf-2021-ac0034>
- [30] Mathieu, C. (2021). *Dark Personalities in the Workplace*. Academic Press.
- [31] McDonagh, A., Giæver, T. H., Mifsud, L., & Milton, J. (2021). Editorial Introduction-Digital Competence in Teacher Education across Europe. *Nordic Journal of Comparative and International Education (NJCIE)*, 5(4), 1-4. <https://doi.org/10.7577/njcie.4604>
- [32] McGarr, O., & Gavaldon, G. (2018). Exploring Spanish pre-service teachers' talk in relation to ICT: balancing different expectations between the university and practicum school. *Technology, Pedagogy and Education*, 27(2), 199-209. <https://doi.org/10.1080/1475939X.2018.1429950>
- [33] McGarr, O., & McDonagh, A. (2021). Exploring the digital competence of pre-service teachers on entry onto an initial teacher education programme in Ireland. *Irish Educational Studies*, 40(1), 115-128. <https://doi.org/10.1080/03323315.2020.1800501>
- [34] Mehta (2022, April 13). Why hybrid classes need to be made permanent. *Oman Daily Observer*. <https://www.omanobserver.om/article/1117708/opinion/international/why-hybrid-classes-need-to-be-made-permanent>
- [35] Milutinović, V. (2022). Examining the influence of pre-service teachers' digital native traits on their technology acceptance: A Serbian

- perspective. *Education and Information Technologies*, 1-29. <https://doi.org/10.1007/s10639-022-10887-y>
- [36] Nasser, R. (2019). Educational reform in Oman: System and structural changes. In *Education systems around the world*. IntechOpen. <http://dx.doi.org/10.5772/intechopen.84913>
- [37] Oman (n.d.). *Countries and their Culture*. <https://www.everyculture.com/No-Sa/Oman.html#ixzz7lZ3sXOxn>
- [38] Oman expanding the use of technology in classrooms of all levels (n.d.). *Oxford Business Group*. <https://oxfordbusinessgroup.com/analysis/vital-tools-authorities-are-expanding-use-technology-classrooms-all-levels>
- [39] Oman has 3.9 million internet users (2020, April 26). *The Arabian Stories* <https://www.thearabianstories.com/2020/04/26/oman-has-3-9-million-internet-users-facebook-top-among-all-other-social-networks/>
- [40] Oman's Ministry signs strategic partnership with Microsoft on digital transformation in education (2020, December 21). *Microsoft News Center* <https://news.microsoft.com/en-xm/2020/12/21/omans-ministry-signs-strategic-partnership-with-microsoft-on-digital-transformation-in-education/>
- [41] Omantel's network upgrades help meet increased demands during pandemic (2020, July 14). *Muscat Daily*. <https://www.muscatdaily.com/2020/07/14/omantels-network-upgrades-help-meet-increased-demands-during-pandemic/>
- [42] Paetsch, J., & Drechsel, B. (2021). Factors Influencing Pre-service Teachers' Intention to Use Digital Learning Materials: A Study Conducted During the COVID-19 Pandemic in Germany. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.733830>
- [43] Polat, M. (2021). Pre-Service Teachers' Digital Literacy Levels, Views on Distance Education and Pre-University School Memories. *International Journal of Progressive Education*, 17(5), 299-314.
- [44] Rosa, J. D. C., Bucheli, M. G. V., Galán, J. G., & Meneses, E. L. (2019). Integrating the digital paradigm in higher education: ICT training and skills of university students in a European context. *International Journal of Educational Excellence*, 5(2), 47-64. <https://doi.org/10.18562/IJEE.048>
- [45] Santos, C., Pedro, N., & Mattar, J. (2021). Digital competence of higher education professors: analysis of academic and institutional factors. *Obra digital*, (21), 69-92. <https://doi.org/10.25029/od.2021.311.21>
- [46] Siddiqui, S. (2019, July 17). What other oil rich states can learn from Oman? *The New Arab*. <https://www.newarab.com/analysis/what-other-oil-rich-states-can-learn-oman>
- [47] Sillat, L. H., Tammets, K., & Laanpere, M. (2021). Digital competence assessment methods in higher education: A systematic literature review. *Education Sciences*, 11(8), 402. <https://doi.org/10.3390/educsci11080402>
- [48] Silva-Quiroz, J. & Morales-Morgado, E.M. (2022). Assessing digital competence and its relationship with the socioeconomic level of Chilean university students. *International Journal Educational Technology Higher Education* 19, 46. <https://doi.org/10.1186/s41239-022-00346-6>
- [49] Study looks at students online. (2021, January 6). *Oman Daily Observer*. <https://www.omanoobserver.com/article/5451/Local/study-looks-at-students-online-english-competence>
- [50] Syahrin, S. (2020). Less Is More: An Implementation of an Extensive Reading Program in an English Proficiency Course in the Sultanate of Oman. *International Journal of English Linguistics*, 10(6). <https://doi.org/10.5539/ijel.v10n6p171>
- [51] The Council of European Professional Informatics Societies (CEPIS) (n.d.). Awareness of the European Digital Competence Framework for citizens (DigComp) framework among the CEPIS community. <https://cepis.org/digcomp-report-2021/>
- [52] The Education Council Sultanate of Oman (2018). *The National Strategy for Education 2040*. <https://www.educouncil.gov.om/downloads/Ts775SPNmXDQ.pdf>
- [53] Tømte, C., Enochsson, A., Buskqvist, U., & Kårstein, A. (2015). Educating online student teachers to master professional digital competence: The TPACK-framework goes online. *Computers & Education*, 84, 26–35. <https://doi.org/10.1016/j.compedu.2015.01.005>
- [54] Touron, J., Martín, D., Navarro, E., Pradas, S., & Inigo, V. (2018). Construct validation of a questionnaire to measure teachers' digital competence (TDC). *Revista española de pedagogía*, 76(269), 25-54. <https://doi.org/10.22550/REP76-1-2018-02>
- [55] Tuzlukova, V., & Usha Prabhukanth, K. (2018). Critical thinking and problem solving skills: English for science foundation program students' perspectives. *Collection of papers of the faculty of philosophy of the University of Pristina, XLVIII* (3), 37.
- [56] UNICEF (n.d.). *Reimagine education*. <https://www.unicef.org/reimagine/education>
- [57] United Nations (2016) SDG-Education 2030 Steering Committee <https://sustainabledevelopment.un.org/index.php?page=view&type=30022&nr=100&menu=3170>
- [58] Wajeha Al-Ani (2017) Alternative education needs in Oman: accommodating learning diversity and meeting market demand, *International Journal of Adolescence and Youth*, 22:3, 322-336. <http://dx.doi.org/10.1080/02673843.2016.1179204>
- [59] Wazzan, M. (2020, April 7). 15 minutes of Fame: Online learning in the Coronavirus era. *Al-Fanar Media*. <https://www.alfanarmedia.org/2020/04/15-minutes-of-fame-onlinelearning-in-the-coronavirus-era/>
- [60] Zhao, Y., Sánchez Gómez, M. C., Pinto Llorente, A. M., & Zhao, L. (2021). Digital competence in higher education: Students' perception and personal factors. *Sustainability*, 13(21), 12184. <https://doi.org/10.3390/su132112184>

# A Survey on Cloudlet Computation Optimization in the Mobile Edge Computing Environment

Layth Muwafaq<sup>1</sup>, Nor K. Noordin<sup>2</sup>, Mohamed Othman<sup>3</sup>, Alyani Ismail<sup>4</sup>, Fazirulhisyam Hashim<sup>5</sup>

Department of Computer and Communication Systems Engineering-Faculty of Engineering,  
Universiti Putra Malaysia, Serdang 43400 Malaysia<sup>1, 2, 4, 5</sup>

Research Centre of Excellence for Wireless and Photonics Network (WiPNET)-Faculty of Engineering,  
Universiti Putra Malaysia, Serdang 43400 Malaysia<sup>2, 5</sup>

Department of Communication Technology and Networks, Universiti Putra Malaysia (UPM), Serdang 43400, Malaysia<sup>3</sup>

Laboratory of Computational Sciences and Mathematical Physics-Institute for Mathematical Research (INSPeM), Universiti  
Putra Malaysia (UPM), Serdang 43400, Malaysia<sup>3</sup>

**Abstract**—Mobile Edge Computing (MEC) uses to perform computation operations at the edge of a network for mobile devices. This allows the deployment of more powerful and efficient computing resources in a cost-effective, lightweight and scalable manner. MEC can optimize mobile device performance, enhance security and privacy, improve battery life, provide increased bandwidth, and reduce latency across wireless networks. Cloudlets are a new concept of computations that can perform at the edge of the networks. The service provider can deploy cloudlets services in a MEC environment with the ability for mobile devices to offload their tasks to cloudlets. In the MEC environment, the offloading problem depends on cloudlets' availability of computation resources. Also, the deployment method of cloudlets in the environment will affect the task offloading. This paper investigates the approach to the cloudlet deployment and task offloading problem in the MEC environment. First demonstrate that the problem has to be considered a Multi-objective optimization problem since it needs more than one objective to be optimized. Then prove that the problem is NP-completeness, give an overview of existing solutions using the meta-heuristic algorithms, and suggest future solutions for this problem. Finally, explain the advantages of using Variable-length of solution space with meta-heuristic algorithms for this problem.

**Keywords**—Mobile edge computing, cloudlet deployment; task offloading; mobile device; multi-objective optimization; meta-heuristics; variable-length

## I. INTRODUCTION

Mobile Edge Computing (MEC) is a technology that enables computation and storage capabilities at the edge of a mobile network, closer to the end-users [1]. It extends cloud computing to the network's edge, allowing for delivering low-latency and high-bandwidth services to mobile users [2]. The end-users typically execute the applications on their resource-constrained mobile devices for current internet-enabled applications requiring fast processing and less response time [3]. MEC is designed to address the challenges of providing low-latency and high-bandwidth services to mobile users, such as those required for augmented reality, virtual reality, and Internet of Things (IoT) applications [4]. Furthermore, the mobile devices at the edge of the network offload their computation tasks to the ( edge-server, edge-cloud, sometimes

referred to as Cloudlet) instead of the remote cloud, which will decrease the response time for offloaded tasks (low latency) and reduce overcrowding in the back-haul networks [5].

Cloudlet is a new computing paradigm introduced to the Mobile Edge Computing (MEC) service framework. It allows computing resources to be closer to mobile devices. The cloudlets location is essential to the delay tolerance of mobile devices, primarily in a large-scale Wireless Metropolitan Area Network (WMAN) that consists of hundred Base Stations (BSs) [6], where mobile devices can access the cloudlets. The capacity of cloudlet is much smaller than cloud computing as edge computing is supplied with one or a few servers due to the limitation of space and cooling requirements [7]. MEC has been adopted to allow mobile devices to offload their tasks to the cloudlets because of the limited processing, small storage and low computational capabilities of mobile devices [8]. The high quality of service (QoS) requirements for the highly interactive applications, which include low latency and high throughput, are computationally demanding [9]. Mobile computing has recently experienced a paradigm change from mobile central cloud computing to MEC, fueled by hopes for 5G and 6G IoT connectivity [10]. In order to support latency-intensive computing applications and critical latency for mobile devices with limited resources, MEC's main purpose is to push mobile computing, network control and storage to the network's edges [11]. In the MEC environment, cloudlets are usually deployed collocated with cellular base stations so that the mobile devices can offload tasks to the nearby cloudlets.

Offloading is a technique used in the MEC environment to increase the effectiveness of mobile device applications by moving resource-intensive activities to nearby cloudlets [12]. Offloading in MEC mostly refers to running resource-intensive applications on behalf of local mobile devices to minimize workloads, overhead, and processing costs compared to local computing. To perform compute offloading, mobile devices and cloudlets must operate offloading frameworks [13]. Many technical articles view the topic as incredibly important to provide new ways of reaching the objectives in the offloading criteria. Most of the strategies presented in these technical publications were based on mathematics, model-based, machine learning, game theory,

heuristic-based, meta-heuristic, or a combination of the abovementioned techniques. In the MEC context, computation offloading problems are a very difficult challenge [14]. The primary drawback of offloading work to a remote cloud is the latency, which disrupts user experiences in interactive applications like mobile gaming [15]. Cloudlets get around this problem by giving users low-latency access to network-edge computing resources, which significantly boosts the efficiency of mobile applications [16]. The primary issue with WMAN is the deployment of fewer cloudlets with good services to end user. From the perspective of network management, it is costly to place a cloudlet at each BS to service end users [17].

Multi-objective optimization techniques are an excellent approach in this situation. In multi-objective optimization, as opposed to single-objective optimization, the search is for a collection of non-dominated solutions known as the Pareto optimal set rather than a single optimal solution, which must be optimized [18]. The non-dominated objective solutions are the ones that provide the best potential compromises between the many objectives of the problem (i.e., these solutions cannot enhance one objective without affecting another). The decision-makers, in this case, the service providers, are given access to such non-dominated solutions so that they may choose the one that caters to their specific demands and requirements in the most effective manner. The computational methods that are currently available to solve multi-objective optimization problems include meta-heuristics and high-level strategies governing underlying techniques. Additionally, in contrast to the conventional mathematical programming approaches used to solve multi-objective optimization problems, meta-heuristics can create many members of the Pareto optimal set in a single iteration [19]. Most researchers consider the problem as a Single Objective Optimization in the related works, while a few deal with it as a Multi-Objective Optimization problem.

A meta-heuristic optimization has shown its effectiveness in tackling several NP-hard problems. Computer science, networking, communication, robotics, and manufacturing, are just a few areas where this mechanism has been used in the real world. The literature on meta-heuristic optimization has yet to adequately address the multi-objective optimization component, despite its prevalence in many problems. The meta-heuristic algorithms have a substantial impact on solving cloudlet deployment and task offloading optimization issues. Consequently, this paper provides a survey of published articles about cloudlet deployment and task offloading to summarize the problems that need solving for future research. First discuss the challenges addressed partially or entirely for the problem and suggest several promising directions for future research to minimize time, reduce cost, and save energy. Second, demonstrate that the problem is multi-objective optimization since it deals with more than one objective to be optimized and proves it is an NP-completeness. Third, explain the diverse current, well-known traditional heuristic and meta-heuristic algorithms for cloudlet

deployment and task offloading problems in the MEC environment, emphasizing meta-heuristics. Fourth present the capabilities of using the variable-length searching approach within meta-heuristic optimization algorithms for solving this type of problem, especially with the constant change in the required number of cloudlets to deploy in MEC based on computing requirements. This work can be helpful for academia and companies regarding service provisioning in the mobile edge computing environment.

The rest of this paper is organized as follows: Section II presents a background and an overview of the relevant literature, which helps to understand the rest of the article. Sections III introduce the optimization technique. Section IV demonstrates the multi-objective optimization and the meta-heuristics algorithms. Section V Provides the Variable-length approach for solution space and its benefits for future research, and finally, the conclusion are provided in Section VI.

## II. BACKGROUND

This section presents an overview of the relevant works, which assists in understanding the rest of the article. Cloudlet deployment and task-offloading (for simple CDTO) have been fertile research ground for several years. Thus, many papers on solutions related to this area are in the literature.

### A. Mobile Edge Computing

One of the essential fields on the technology scene today is Cloud Computing, which is one of the elements preparing for the future. Thanks to advancements in wireless communication and mobile computing, in the last decade, new smartphone device services have prospered in diverse fields such as transportation, mobile payment, and social media [20]. The spread of mobile devices and their constant presence in daily life has generated massive traffic between end users and remote clouds [21]. To plan for the increasing data traffic in the following years and the need for low-latency computation resources near the users, network service providers are gradually turning to "Mobile Edge Computing" to bring cloud computing capabilities to the edge of the network [22]. Users at the edge of the network offload their computation tasks to the ( edge-server, edge-cloud, sometimes referred to as Cloudlet) instead of remote clouds, which will decrease the response time for offloaded tasks (low latency) and reduce overcrowding in the back-haul network [20].

Due to the limitations of the cloud environment (i.e., long-distance from mobile devices and limited geographical distribution), mobile devices with hungry applications for computation resources need to offload tasks to the nearby clouds closer to the end-devices through newer computing paradigms, such as cloudlets in Mobile Edge Computing environment (MECE) [23]. Cloudlets are a distributed decentralized infrastructure with nearby mobile devices that can leverage the computing and storage resources of the cloudlets. Table I presents the comparative summary of Cloud and Cloudlet.

TABLE I. COMPARISON BETWEEN CLOUD AND CLOUDLET

Feature	Cloud	Cloudlet
Cost	High	Low
End-to-end latency	High	Low
Infrastructure	Centralized	Decentralized
Deployment Environment	Large Data Center	Can be deployed anywhere
Bandwidth for end-user	Low	High
Offline Availability	Not available	Available
Computing Power	High	Low
Resource Elasticity	High	High
Availability	High	High
Access to resources	through core network	typically via 1-hop wireless gateway
Resources at individual locations	Many	Few
Geo-distribution of computing resources	locally clustered	widespread
Offloading granularity	mostly entire applications	computationally intensive and latency-critical parts
User Experience	Satisfactory QoE	Excellent QoE
Network resource sharing at individual locations	Large number of users	Limited number of users

### B. Wireless Metropolitan Area Networks

The wireless metropolitan area networks (WMANs) have emerged in recent years as a public network that enables mobile devices to easily access the abundant computing resources in urban cities as the number of mobile devices continues to grow [24]. In addition, wireless broadband connectivity requirements are growing (i.e., 5G and 6G), and the technology continues to evolve to meet these requirements [25]. It is currently evolving as a public network to expand mobile device performance, covering the metropolis and enabling mobile devices to access abundant computing resources [26]. Compared to Wireless Local Area Networks (WLANs), Wireless Metropolitan Area Networks (WMANs) offer wireless communications over significantly broader geographic regions. The problem of CDTO gets significantly worse when considering the deployment of cloudlets in WMANs because of the size of the WMANs [27]. Due to the high population density in urban regions, many users will have access to cloudlets. As a result, cloudlets will be more cost-effective since they will be less likely to stay idle. Furthermore, due to the population in WMANs, service providers can take advantage of economies by deploying a small number of cloudlets to provide services through the WMANs [20] and making cloudlets accessible to the general public.

### C. Cloudlet-deployment and Task-offloading (CDTO)

Cloudlets deployment is the first step in implementing mobile edge computing. Service providers always hope to provide services to as many users as possible under limited funds. The idea of treating cloudlets as a separate “data center in a box” must be scrapped. There are obvious benefits of connecting several cloudlets to form a network of cloudlets that can be distributed in the wireless metropolitan area networks, collocated with the base stations [20]. Cloudlet can be a single server or a cluster of servers, collocated with base stations, and mobile devices can access cloudlets close to the

range [28]. Cloudlets can also be considered as offloading mobile device destinations, aiming to save energy [29] and reduce the latency between the remote clouds and mobile devices [30]. The close physical proximity between cloudlets and mobile devices is a significant benefit of cloudlets over the remote cloud. So that allows for low latency in communication, thus improving the QoS and the user experience of interactive applications [31]. To overcome the restrictions of mobile devices in the offloading process technique, mobile devices offload their tasks to the MECE [32]. Task-offloading improves the computing efficiency of mobile devices while reducing latency and overall spent energy if adequately designed and planned [5]. There are four types of delays in task-offloading from mobile devices to cloudlets: uploading, queuing, execution, and downloading the result. A typical process flow of task-offloading in WMAN is illustrated in Fig. 1.

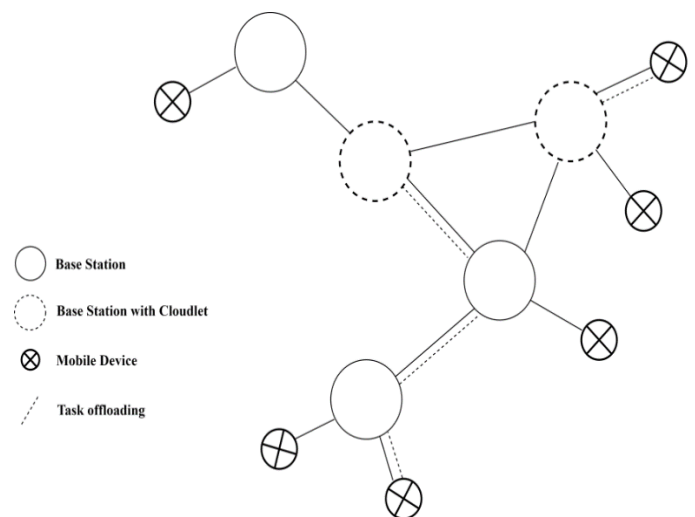


Fig. 1. Process flow of the task-offloading in WMAN.

In related works, there are mainly three metrics expressing system requirements that can be optimized, the task response time [33], the cost of service for service provider and user side [34], and the energy consumption (the energy consumed by the device while task-offloading and the power needed by cloudlet to process tasks) [35]. In addition, some issues have been partially addressed in previous works, which are reliability, load balancing, cache content, task migration, user mobility, and cloudlet mobility. The exciting issues for related works that dealt with the problem as a single objective optimization are shown in Table II.

### 1) Response Time

The QoS metric in the design of offloading methods is the response time, which is the time between offloading the task from a mobile device to a cloudlet, processing it, and receiving the result back [36]. Furthermore, the response time is also defined as the delay [37], or latency [38] in some related works. Many aspects influence the response time of a task as followings:

- The communication delay between the mobile device and the base station while offloading the task to the cloudlet.
- The communication delay between the base station and the attached cloudlet, which is always minimal while using fiber optics cables.
- The communication delay while routing the offloaded task from overloaded to under-loaded cloudlet.
- The queuing time before processing in the cloudlet.
- The processing time on the cloudlet.
- The communication delay between the base station and mobile device while receiving the result.

Fig. 2 shows the task lifetime, from the mobile device to cloudlet, which is an uplink delay, queuing and execution at cloudlet, which is a processing delay, and from cloudlet to mobile device, which is a downlink delay.

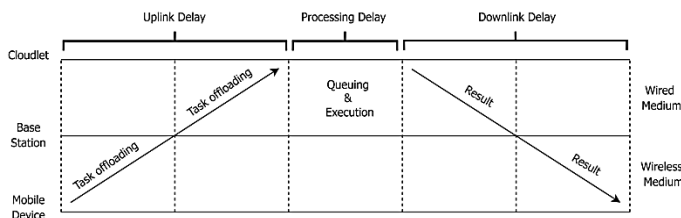


Fig. 2. Task service delay.

### 2) Cost

The cost can be categorized into two sides: service providers and users. For the service provider, the cost of deploying cloudlets is related to two factors [21], which are site rentals and computations needs. The former means that the more locations are selected to deploy cloudlets, the higher the cost. The latter aspect tells us that the greater the computation needs, the greater the number of cloudlets, resulting in higher costs. On the other hand, users look for the best service at a reasonable price. The service price (the

amount user have to pay) mainly specifies the willingness of users to pay for services [39].

### 3) Energy

Most existing works on MECE mainly focus on saving energy for mobile devices by offloading tasks to the cloudlets. Due to the battery life limitation for various mobile devices, the energy consumed by the mobile device processing the tasks is one of the most significant issues [40]. The energy is mainly consumed by mobile device data processing and transmission. Due to the limitation of processing in a mobile device [39], it is better to offload tasks from mobile devices to nearby cloudlets for energy-saving. The offloading of a task can decrease the processing energy by reducing the processed data size but increases the transmission energy by increasing the size of transmitted data [35]. On the other hand, service providers are looking for energy savings for cloudlets, reducing power costs.

### 4) Reliability

Reliability is a significant component that influences user QoS and service provider profit [41]. The significant causes of reliability decreases are failures or errors in the mobile edge computing environments. Failures and errors in the system can occur during task-offloading, processing, routing tasks from one cloudlet to the other, and receiving the result back [42]. Improving the reliability of task-offloading under the resource capabilities of the cloudlets and the dynamic network is an important challenge in the MECE.

### 5) Load Balancing

The cloudlets contain single or multiple servers to process the offloaded tasks. The load inequality in these cloudlets leads to different response times for offloaded tasks. Therefore, the load balancing between cloudlets for task-offloading becomes essential [22]. Increasing the number of servers in the cloudlet makes balancing user workload less critical as the cloudlets can handle higher loads [43]. On the other hand, increasing the number of users will increase the task waiting time. Once the cloudlet reaches its maximum workload, it must route its overflow users' requests to the other cloudlets for better performance [44]. This means that a better strategy is to assign user tasks to under-loaded cloudlets rather than keeping them in an overloaded cloudlet [22]. Hence, it is clear that balancing the workload between cloudlets is necessary by mapping the tasks of users to different cloudlets.

### 6) Cache Content

When a task is offloaded to a cloudlet, the critical impact on task execution performance is data transmission delay. One of the most successful solutions is to cache data for the offloaded task in advance. Furthermore, because user data access patterns in WMANs are challenging to forecast [45], creating an adequate data caching technique with high access hit ratio is challenging. The number of content cached at cloudlet increases with the number of offloaded tasks increasing [46]. Content cache performance significantly impacts delay and bandwidth [47]. For example, the popular cached content in cloudlet can reduce the delay and the bandwidth usage of subsequent access to the same content.



7) Task Migration

In the MECE, the inter-flow between cloudlets is essential to make a load-balancing and support user mobility. In the network of cloudlets, to make load-balancing, tasks have to be forwarded from the overloaded cloudlet to the under-loaded cloudlet, so there will be no idle cloudlet in the system [48]. Hence, as shown in Fig. 3, to support user mobility in the system, tasks must route from one cloudlet to another cloudlet so that the task will be close to the mobile user for result download [49].

8) User Mobility

To support user mobility, edge computing requires the flexible and scalable deployment of cloudlets for the inherent dynamism of the operating environment and various applications, some of which require real-time response [50]. In related works, WMAN's user mobility pattern has attracted very little attention. A mobile user may move around the area covered by WMAN with specific transmission patterns, but different movement patterns may change the traffic size of the network. The cloudlet placement and task offloading strategy should reflect the mobility pattern better to meet the requirements of mobile devices to access network resources [51]. The mobility pattern is reflected by the connection possibilities between mobile devices and BS. Furthermore, the mobility pattern of mobile devices is highly correlated with service access delay. Connection failures may happen during user mobility with the poor quality of the wireless connection between mobile devices and BSs [52]. However, the next generation of cellular communication has overcome this issue. The excellent service would necessitate frequent wireless handover between multiple BSs to guarantee user QoS during

movement. For constant QoS, mobile devices may be serviced by many cloudlets. Because of user mobility, wireless handover and service migration may be performed frequently, putting a heavy load on the other network entities [53].

To support user mobility, task migrations have to perform when a mobile user moves from the service region of one cloudlet to another. As mentioned four types of task service delay; there will also be a delay in task migration during user movement. Fig. 3 illustrates the task migration during user movement. To provide QoS for users, mobility management should make a wireless handover decision to select optimum BSs and cloudlets [54]. The primary goal of mobility management is to provide mobile computing services to continuous and uninterrupted customers during user movements.

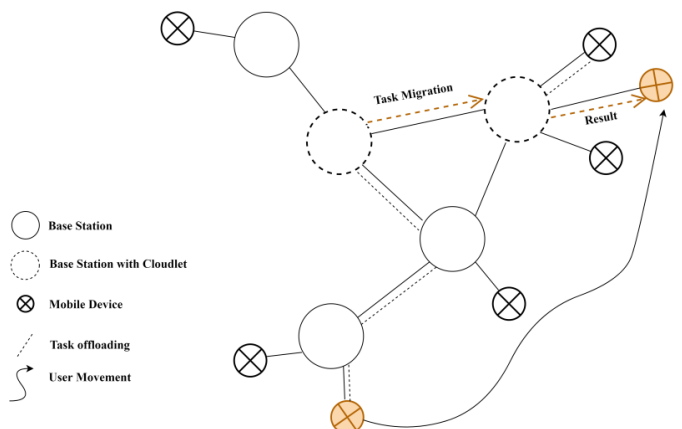


Fig. 3. Task migration while user movement.

TABLE II. RELATED WORKS CONSIDERING A SINGLE OBJECTIVE OPTIMIZATION

Authors	Cloudlet Deployment	Task offloading	Time	Cost	Energy	Reliability	Load balancing	Cache content	Task Migration	User Mobility	Cloudlet Mobility
[20]	√	√	√	×	×	×	√	×	×	×	×
[55]	√	×	√	×	×	×	×	×	×	×	×
[37]	√	×	√	×	×	√	√	×	×	×	×
[56]	×	√	√	×	√	×	√	×	×	×	×
[31]	√	×	√	√	×	×	×	×	×	×	×
[49]	×	√	√	√	×	×	×	×	√	√	×
[7]	√	×	√	√	×	×	×	×	√	√	×
[57]	√	×	√	×	×	×	×	×	√	√	×
[58]	√	×	√	√	√	×	×	×	×	×	×
[59]	√	×	√	×	√	×	×	×	×	×	×
[60]	√	√	√	×	√	×	×	×	×	×	×
[41]	√	×	√	√	×	√	×	×	×	×	×
[22]	×	√	√	×	×	×	√	√	√	×	×
[61]	√	×	√	×	×	×	√	×	×	√	√
[62]	×	√	√	×	√	×	×	√	×	×	×
[63]	×	√	√	√	√	×	×	×	×	×	×
[42]	×	√	√	×	×	√	×	×	×	×	×
[44]	√	√	√	×	×	×	√	×	×	×	×
[64]	√	√	√	√	×	×	×	×	×	×	×
[19]	√	√	√	×	×	×	×	×	×	×	×
[65]	√	×	√	√	×	×	×	×	×	×	×
[66]	√	×	√	√	√	×	×	×	×	×	×

### 9) Cloudlet Mobility

The cloudlet in the MECE can be static or mobile [12]. In mobile cloudlet networks, it is crucial to figure out how to make load-balancing between all mobile cloudlets so that all resources are utilized and tasks can be processed concurrently and sustainably by different cloudlets, therefore, reducing the average task response time and energy [61]. Cloudlet mobility is essential in MECE to provide good services for users, especially when the population moves (e.g. festivals, sporting events, etc.).

In Table II, all researchers consider the problem as a single-objective optimization. Most of them take the problem on one side (cloudlet-deployment or task-offloading), while very few consider both themes. Furthermore, most existing works mainly focus on minimizing time, reducing cost, saving energy, and load balancing, while the others take little attention. In Table II, various algorithms and methods were employed to address the issues, including greedy algorithms, clustering algorithms, heuristics, meta-heuristics, reinforcement learning, and mathematical programming. Comparisons were made between meta-heuristic and heuristic algorithms, meta-heuristic and greedy algorithms, reinforcement learning and mathematical programming. In addition, research studies such as [44] have compared meta-heuristic and unsupervised learning algorithms, while [65] has compared meta-heuristic and mathematical programming, with results consistently demonstrating the superiority of meta-heuristic algorithms.

### D. Related Survey Works

CDTO has been a fertile study area in recent years. As a result, many surveys and studies in the literature on solutions related to this area. In the work of [67], the problems of mobile cloud computing are presented and provide the most recent mobile cloudlet architecture. Furthermore, propose a hierarchical taxonomy to classify the most recent cloudlet solutions and discover the cloudlet application areas. In addition, it presents the aspects of cloudlet management, like cloudlet discovery, resource management, data security, mobility, and application offloading. The work of [68], studied the fundamental concepts of cloud and edge computing. Presented the application domain classified the state-of-the-art edge computing (Mobile Edge Computing, Cloudlet, and Fog) and the application domain area services such as real-time applications, resource management, data analytics, and security. Furthermore, state the essential requirements that must be achieved for edge computing to be enabled. The work of [69] proposes a taxonomy of task-offloading in edge-cloud environments to examine and categorize related research papers and outline the challenges that still need to be studied before using edge-cloud computing to improve services. To identify the modern processes, the work of [70] provides a review of the machine learning-based computation offloading strategies in the MECE in classical taxonomy. Furthermore, it investigated various strategies and novel approaches related to machine learning-based offloading mechanisms in the MECE ecosystem. Also, taxonomy for classifying various principles of machine learning-based offloading mechanisms was proposed.

Finally, there are many other surveys, each considering a particular aspect of Mobile Edge Computing, which also lacks considering the Multi-objective optimization approach for the problem of CDTO.

## III. OPTIMIZATION

Optimization is a method of finding and comparing appropriate solutions until cannot find any better solutions. Optimization often involves minimizing or maximizing the objective functions. Optimization, in other terms, refers to a collection of methods that may be applied to a mathematical model of the problem. The optimization algorithm provides systematic and efficient methods for producing and comparing new solutions to achieve the optimal solution [71]. Optimization is experimenting with various input-output combinations to find the resulting outputs. Its methods are far more advanced than those used in the computation. Optimization theory has custom-made algorithms to identify the best solution with little processing by utilizing model information. Optimum selection evaluations, suitable trade-offs, and non-intuitive analysis of optimization techniques are utilized to develop a better and faster optimal design [72]. The role of modeling in optimization involves testing out many different combinations of inputs to determine what the output (i.e., the number you are trying to minimize or maximize) will be under each of those circumstances, with many inputs, it is usually not practical to guess and check on an existing system [73], optimization requires a mathematical model of the system, which is just the math that relates the inputs to the outputs. By using a system model to iterate, there is no limit to the number of combinations you can try (maybe just computational limitations). Optimization algorithms are much more sophisticated than guessing and checking a range of varieties (a process known as enumeration) [74]. Fig. 4 outlines the steps usually involved in an optimal design formulation.

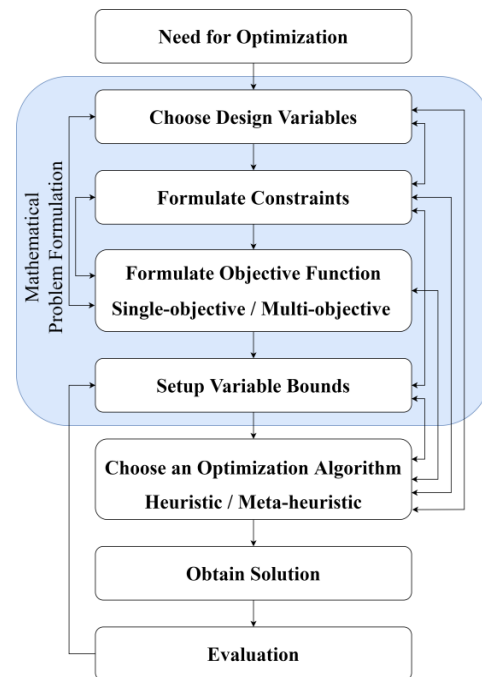


Fig. 4. Steps of problem optimization.

In the Mobile Edge Computing environment, users and the service provider have various requirements. The deployment of a small number of cloudlets in the WMANs to serve many users is significant for service providers. In addition, user satisfaction is one of the most important quality-of-experience (QoE) evaluation measures, especially affecting the service provider's profit due to the penalty when there is any infringement of the Service Level Agreement (SLA), which establishes an agreement between service providers and their users [75]. It significantly influences users' willingness to pay for services in the future. Thus, when offering services, user requirements must be considered and satisfied [40]. Furthermore, in the problem of CDTO, most existing works consider single-objective optimization, while very few consider it a Multi-objective optimization. The relationship "better than" between two solutions is insignificant when dealing with single-objective optimization problems. Whether a function has to be maximized or minimized, the one with the lower or higher fitness value is preferred. The relationship "better than" has to be redefined while dealing with the Multi-objective optimizations. Furthermore, when two non-dominated solutions are compared, there is only one way to rate the best if enforcing specific user preferences. As a result, a new kind of fitness measurement is required [74].

#### IV. MULTI-OBJECTIVE OPTIMIZATION

Problems involve more than one objective in the real world, and conflict naturally leads to trade-off solutions. It is hard for the mobile edge computing service provider to determine the constraints. A feasible solution cannot be found if the constraints are not appropriate. Only one objective is confirmed, and only one solution may be generated if a constrained single-objective optimization problem is considered rather than a multi-objective optimization problem which may not result in a satisfactory trade-off between all objectives [76]. Most researchers in related works dealt with the problem as a single-objective optimization. In contrast, few researchers are dealt with the problem as multi-objective optimization, as shown in Table III.

For the Multi-objective optimization problem arising from Cloudlet-deployment and Task-offloading (for simple MO-CDTO) in MECE, not only the minimization of task completion time and the energy consumption is required, but also the minimization of cloudlet-deployment cost is demanded [60]. In Table III, researchers consider the problem

as multi-objective optimization. Most of them take the problem on one side, which is task offloading. Furthermore, time, cost, energy, and task migration have more extensive attention than others. The algorithms and methods were employed to address the multi-objective optimization problems, including meta-heuristic algorithms, a hybrid of meta-heuristic algorithms, machine learning and reinforcement learning. The results consistently demonstrate the superiority of the hybrid of meta-heuristic algorithms to address the multi-objective optimization problems.

##### A. The NP-Completeness of MO-CDTO Problem

The NP-complete problem is the hardest to solve. It is a problem that if there is an algorithm that can solve any of these problems, then it can solve all of them. Nobody has been able to prove that NP-complete problems are intractable. To prove that the MO-CDTO is an NP-completeness, first must explain that the problem belongs to the NP class and then show it is NP-hardness. When presenting a solution for the problem of MO-CDTO, the problem takes  $O(ML)$  time to evaluate if the total task-offloading probability for each user is no more than one and the maximum workload restriction for each cloudlet. Furthermore,  $O(L)$  time is required to verify that the number of deployed cloudlets at each base station is not greater than one. As a result, verifying a solution for MO-CDTO takes polynomial time. Therefore, MO-CDTO is under the class of NP problems. Based on the explanation of the p-median problem in [81]. When  $\forall 1 \leq i \leq$  the number of users, the computing capacity of user  $i$  is zero, the computing capacity of each cloudlet is infinity, the cost of each cloudlet is equal to zero, the energy consumption of users is constant, the cloudlet deployment cost can be neglected, and the maximum number of cloudlets that may be deployed must be deployed to reduce user task response time. Therefore, the problem of MO-CDTO becomes selecting BSs to deploy the number of cloudlets and assigning each user to a cloudlet so that the sum of the connection delay is minimized from all users to the cloudlets they are assigned. To that end, each user must be assigned to a cloudlet with the smallest communication delay so that the MO-CDTO problem is NP-hard. Furthermore, based on the definition of NP-complete in [82] and [83] "If the problem  $X \in$  NP class and  $X$  is NP-hard so  $X$  is NP-complete", while the problem of MO-CDTO belongs to the NP class and it is NP-hard, so it is NP-complete.

TABLE III. RELATED WORKS CONSIDERING A MULTI-OBJECTIVE OPTIMIZATION

Authors	Cloudlet Deployment	Task offloading	Time	Cost	Energy	Reliability	Load balancing	Cache content	Task Migration	User Mobility	Cloudlet Mobility
[39]	×	√	√	√	√	×	×	×	√	×	×
[29]	×	√	√	×	√	×	×	×	√	×	×
[26]	×	√	√	×	√	×	√	×	√	×	×
[77]	×	√	√	×	√	×	×	×	×	×	×
[78]	×	√	√	√	√	×	×	×	×	×	×
[79]	×	√	√	×	√	×	×	×	√	×	×
[5]	×	√	√	√	√	×	×	×	×	×	×
[76]	√	√	√	√	√	×	×	×	√	×	×
[80]	√	×	√	×	×	×	√	×	×	×	√

**B. Meta-heuristic Algorithms**

The meta-heuristic algorithms are a class of local search algorithms used to solve multi-objective optimization problems. The main objective of these algorithms is to find the optimal solution for the given set of problems based on a given set of constraints. The main idea behind this method is that it considers all the constraints and variables that affect the solution. Users can specify these constraints or automatically generate them by the algorithm. In mathematical programming, heuristic algorithms are used to design solutions to problems as quickly as possible. It may not produce the best solution, but it will give a near-optimal solution in a short time [71]. Some trade-off conditions give an initial idea of whether a heuristic algorithm is a good fit or not for a given problem. One of these conditions is the

completeness of the problem. If several solutions exist for a given problem, it is better to use a meta-heuristic algorithm [74]. A heuristic algorithm generally provides one solution that may not be the best among all the available solutions. Furthermore, meta-heuristic-based optimization has proved effective in solving many NP-hard problems. This method has been used in various real-world applications, including computer science, networking, communication, robotics, and manufacturing [84]. The main difference is that heuristic is the problem-specific method, while meta-heuristic is the problem-independent method that can be applied to many problems. Table IV classifies the heuristic and meta-heuristic algorithms. Meta-heuristic algorithms are divided into nature-inspired and non-nature-inspired.

TABLE IV. CLASSIFICATION OF HEURISTICS AND META-HEURISTICS ALGORITHMS

Heuristics	Meta-heuristics	
	Nature-inspired	Non-nature-inspired
<ul style="list-style-type: none"> <li>• Greedy Algorithms</li> <li>• Best-First Search</li> <li>• Brute-force Search</li> <li>• A* Search</li> <li>• Hill Climbing</li> <li>• Bidirectional Search</li> <li>• Beam Search</li> </ul>	<ul style="list-style-type: none"> <li>• Swarm Intelligence                             <ul style="list-style-type: none"> <li>• Whale Optimization (WO)</li> <li>• Cuckoo search algorithm (CSA)</li> <li>• Particle Swarm Optimization (PSO)</li> <li>• Ant Colony Optimization (ACO)</li> </ul> </li> <li>• Estimation of Distribution Algorithm (EDA)</li> <li>• Simulated annealing (SA)</li> <li>• Evolutionary Computation (EC)                             <ul style="list-style-type: none"> <li>• Differential Evolution (DE)</li> <li>• Evolution Strategy (ES)</li> <li>• Genetic Algorithm (GA)</li> <li>• Genetic Programming (GP)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Guided Local Search (GLS)</li> <li>• Greedy Randomized Adaptive Search Procedure (GRASP)</li> <li>• Iterated Local Search (ILS)</li> <li>• Path Relinking (PR)</li> <li>• Scatter Search (SS)</li> <li>• Tabu Search (TS)</li> <li>• Variable Neighborhood Search (VNS)</li> </ul>

Meta-heuristic algorithms are generated based on the mathematical models of different biological processes [3] and activities [85] that appear in nature, modified and applied according to a specific problem so that the result statistical data can be analyzed. Meta-heuristic algorithms can be used for CDTO performance metrics, such as cost, time and energy consumption, to reach the optimal solutions for the problem of MO-CDTO. The frameworks and procedures for the different evolutionary, swarm and hybrid algorithms are comparable [86]. However, the methods for population initiating, evaluating the initial fitness (the quality of the solution), the strategies for coming up with new solutions, and the iterative procedures are often different [3]. Because of their success in solving complex and important computing problems, meta-heuristic algorithms have achieved widespread during the last two decades [87]. Due to their applicable independence in problem-solving, meta-heuristic algorithms are useful and well respected for solving problems in different fields with highly acceptable performance. Meta-heuristic algorithms are frequently exploited as an efficient way to address NP-hard optimization problems [88]. The most well-known and extensively used under meta-heuristics are evolutionary algorithms (EAs). Individuals, or potential solutions in EAs, are made up of a chromosome (a representation of the problem's variables) and fitness (the quality of solutions) [89]. Individuals are grouped to form populations. Every time a new solution is developed, it is examined to determine its fitness value. A generation occurs when a new one replaces the

present population [90]. Evolution is iterating through consecutive generations that end when a termination condition is met. The most meta-heuristic algorithms used in the related works are the following:

*1) Genetic Algorithm (GA)*

Genetic Algorithms are natural search algorithms inspired by Darwin's theory of evolution. JH Holland first introduced it in 1973. Genetic algorithms can create high-quality solutions for various problems, including search and optimization, by imitating the processes of natural selection, reproduction, and mutation [91]. The population is updated using the traditional GA update approach, which includes binary tournament selection, a two-point crossover operator, and a mutation operator. Furthermore, the elite individuals (those with the highest fitness values in the population) are entirely replicated in the next generation using the elitist preservation mechanism of this algorithm [8].

*2) Simulated Annealing (SA)*

Simulated annealing is a technique for addressing unconstrained and bound-constrained optimization problems. It was first introduced by [92] in 1983. Simulated annealing investigates alternative configurations to produce better solutions to the massive combinatorial problem. The configurations are updated and compared at each iteration, allowing the best configuration to be chosen [93]. Furthermore, a simulated annealing algorithm initially allows other random movements in the neighboring search space, but

the activities are reduced over time. The algorithm chooses neighboring states with a cost lower than the current state or the exact cost [94].

### 3) Whale Optimization (WO)

Whale Optimization is a new nature-inspired meta-heuristic optimization algorithm that simulates humpback whale social behavior. The bubble-net hunting approach inspired the algorithm. Mirjalili and Lewis, two Australian researchers, introduced the Whale Optimization Algorithm in 2016 [95]. The WOA offers a lot of potential benefits. It is unaffected by early solutions, which can considerably impact some traditional algorithms. It also contains adaptive mechanisms to balance explorative and exploitative behaviors appropriately. It is widely used in various domains, including engineering optimization, feature selection, and parameter extraction [96].

### 4) Particle Swarm Optimization (PSO)

The particle swarm optimization algorithm is based on the study of bird predation behavior. Its straightforward premise is to discover the best solution through collaboration and sharing information between individuals in the swarm. The PSO is one of the meta-heuristic algorithms proposed by [97] in 1995 based on the social swarm evolving notion. According to experimental studies and applications, the PSO is highly competitive in optimization. The essential idea of particle swarms is the social interaction that results in collective intelligent behavior. Both the set of entities and the swarm of birds move with two velocity components. The first one attracts them to the best global entity, and the second attracts them to the best local entity [84].

### 5) Ant Colony Optimization (ACO)

In the ACO technique, natural ant behavior is crucial in determining the optimum route between colonies and food sources. This concept was first presented as the "ant system" [98] in 1992. As they go forward, the ants expel the pheromones. Pheromones form the shortest paths over time, and the intensity of the pheromone helps to identify the quickest route to the food source. Indeed, the ACO was inspired by the ant behavior to determine the shortest path between anthill and the location of the food supply [99].

### 6) Cuckoo Search Algorithm (CSA)

Cuckoo Search is a meta-heuristic optimization algorithm used to solve optimization problems. Yang and Deb invented it in 2009 [100]. The CSA is a nature-inspired meta-heuristic algorithm based on cuckoo brood parasitism, and Levy flights random walks. The CSA is based on the parasitism of the offspring of a bird species known as the cuckoo. The CSA algorithm implements the Lévy flights [100]. Usually, the parameters of the cuckoo search are kept constant for a particular duration; this results in a decrease in the algorithm's efficiency.

### 7) Differential Evolution (DE)

Differential evolution (DE) is one of the most well-known generation EAs, developed by Storn and Price [101] in 1997 and effective for constrained optimization problems in nonlinear and multimodal environments. Differential Evolution is a generic Nature-inspired population-based

global-search meta-heuristic optimization algorithm. Therefore, it is largely suited for numerical optimization problems based on vector differences.

### 8) Tabu Search (TS)

Tabu search (TS) is a heuristic method proposed initially by Glover in 1986 [102]. Tabu search is a potential tool to discover a feasible optimal solution from a limited set of solutions. In other words, it is an optimization approach that uses a guided local search procedure that avoids local optimum and denies moves to points already visited in the search space using the so-called tabu list [102].

The common trend of EAs is hybrids of different evolutionary algorithms. These hybrids are relatively easy since most EAs can use the same population table and gene presentation. Hybrids are mainly accomplished to avoid the shortcomings of one algorithm; for example, using a hybrid of two algorithms, one for a local search and the other for a global search.

## V. VARIABLE-LENGTH

While there is no proper definition for a vector of a variable-length searching of solution space, therefore, most optimization theories depend on the fixed-length assumption to represent solutions [73]. There are several cases of variable-length approaches in which the number of variables is not fixed. One of these cases is CDTO. Standard optimization procedures may be used by considering a fixed number of variables. However, a sub-optimal length will result in a sub-optimal solution [86]. The algorithms can be executed iteratively and change the given length until an optimal solution is found. However, this is inefficient and impractical if there is an extensive range of possible lengths. For other fields of study, Table V shows the algorithms developed in meta-heuristic optimization with the variable length features. Variable-length algorithms are better since their solution vectors can vary in length [103].

Overall, as shown in Table V of the reviewed algorithms, the majority of them were developed to support a single objective except for the work of [89], which is based on an evolutionary algorithm and was applied only to a Bi-objectives problem without investigating the proposed algorithm and the work of [84], which suffers from weak interaction between solutions.

The traditional meta-heuristic optimization algorithms consider a fixed length of solution space, but these solutions do not apply to many real-world problems. The specific values of some decision variables might generate another or disable others, which cause the variable-length nature of solution space because of the different lengths of solutions. Multi-objective nature comes from having more than one objective to be optimized, e.g., delay, cost, and energy. Hence, MO-CDTO is a multi-objective variable-length optimization problem; the variable-length nature comes from the different number of cloudlets to deploy according to the computing requirements. Furthermore, the related works for the problem of MO-CDTO lack a single or multi-objective optimization algorithm with the variable-length feature.

TABLE V. THE PRESENT VARIABLE-LENGTH OPTIMIZATION FOR DIFFERENT FIELDS

Article	Application	Decision space	Algorithm	Limitation	Number Of objectives
[104]	Path planning	Path points	Genetic algorithm	Single objective	1
[73]	Laminate stacking Wind farm Sensor coverage	Based on the problem	Metameric Genetic	Single objective	1
[105]	Changing the topology of Convolutional Neural Networks	Encoding neuron in the layer	Particle swarm optimization	Single objective	1
[89]	Laminate stacking problem / angle-based transformation	Vector of angles on the plies of a laminate / general mathematical form	Multi-objective evolutionary algorithm	the application of the proposed algorithm is not investigated	2 / 3
[86]	A coverage and a wind farm problem	Based on the problem	Evolutionary algorithm	Single objective	1
[106]	Wireless Sensor Network deployment	Sensors location and coverage area	Genetic algorithm	Single objective	1
[84]	Wireless Sensor Network deployment	Sensors location and coverage area	Particle swarm optimization	Weakness of Inter-class interaction	2
[107]	Sensor node scheduling	Deciding the sensor that will send data	Genetic algorithm	Single objective	1

## VI. CONCLUSION

Mobile edge computing has become an important technology to overcome some of the inherent constraints of mobile devices. Service providers are looking to provide excellent and cheap service to the MECE end-users. Furthermore, researchers focus on three main things in the MECE, which are minimizing time, reducing cost, and saving energy while solving different issues. Hence, the issues are related to each other. In contrast, the reliability and load-balancing affect time and energy, the cached content affects time, and the task migration and cloudlet mobility support the user mobility for good service and reducing time. The problem of CDTO has become an exciting research area. It has to be considered a multi-objective optimization problem since it needs more than one objective to be optimized. The problem of MO-CDTO is NP-complete because it belongs to the NP class and is NP-hard. The Meta-Heuristic algorithms have to be used for this type of problem as it is practical to solve MOO problems. For feature works, variable length of solution space is an appropriate approach for the problem of MO-CDTO because of the variation in the number of cloudlets to be deployed. A hybrid of two meta-heuristic algorithms with the variable-length aspect may generate powerful solutions. Many issues still need to address in the MECE, like reliability, user mobility, and load-balancing between cloudlets. This study is excellent material for future researchers to have an overview of the problem and take the research forward to resolve the unaddressed issues.

## REFERENCES

[1] X. Xu, X. Liu, X. Yin, S. Wang, Q. Qi, and L. Qi, "Privacy-aware offloading for training tasks of generative adversarial network in edge computing," *Information Sciences*, vol. 532, pp. 1–15, 2020, doi: 10.1016/j.ins.2020.04.026.

[2] M. T. Kabir and C. Masouros, "A Scalable Energy vs. Latency Trade-Off in Full-Duplex Mobile Edge Computing Systems," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5848–5861, 2019, doi: 10.1109/tcomm.2019.2915833.

[3] E. H. Houssein, A. G. Gad, Y. M. Wazery, and P. N. Suganthan, "Task Scheduling in Cloud Computing based on Meta-heuristics: Review, Taxonomy, Open Challenges, and Future Trends," *Swarm and Evolutionary Computation*, vol. 62, no. October 2020, 2021, doi: 10.1016/j.swevo.2021.100841.

[4] X. Zhao, Y. Shi, and S. Chen, "MAESP: Mobility aware edge service placement in mobile edge networks," *Computer Networks*, vol. 182, no. April, p. 107435, 2020, doi: 10.1016/j.comnet.2020.107435.

[5] P. Wang, K. Li, B. Xiao, and K. Li, "Multi-objective Optimization for Joint Task Offloading, Power Assignment, and Resource Allocation in Mobile Edge Computing," *IEEE Internet of Things Journal*, vol. 4662, no. c, pp. 1–12, 2021, doi: 10.1109/JIOT.2021.3132080.

[6] Z. Xu, W. Liang, W. Xu, M. Jia, and S. Guo, "Capacitated cloudlet placements in Wireless Metropolitan Area Networks," in *Proceedings - Conference on Local Computer Networks, LCN, 2017*, vol. 26-29-Octo, pp. 570–578, doi: 10.1109/LCN.2015.7366372.

[7] X. Guan, X. Wan, T. Wang, and Y. Li, "A long-term cost-oriented cloudlet planning method in wireless metropolitan area networks," *Electronics (Switzerland)*, vol. 8, no. 11, p. 1216, 2019, doi: 10.3390/electronics8111213.

[8] X. Zhao, C. Lin, and J. Zhang, "Cloudlet deployment for workflow applications in a mobile edge computing-wireless metropolitan area network," *Peer-to-Peer Networking and Applications*, vol. 15, no. 1, pp. 739–750, 2022, doi: 10.1007/s12083-021-01279-z.

[9] M. Hui, J. Chen, Y. Zhou, B. He, K. Wu, and L. Yang, "Server Deployment and Load Balancing in Stochastic Mobile Edge Computing Networks," *IEEE Communications Letters*, vol. 26, no. 5, pp. 1194–1198, 2022, doi: 10.1109/LCOMM.2022.3151467.

[10] S. Lai, R. Zhao, S. Tang, J. Xia, F. Zhou, and L. Fan, "Intelligent secure mobile edge computing for beyond 5G wireless networks," *Physical Communication*, vol. 45, p. 101283, 2021, doi: 10.1016/j.phycom.2021.101283.

[11] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017, doi: 10.1109/COMST.2017.2745201.

[12] X. Jin, F. Gao, Z. Wang, and Y. Chen, "Optimal deployment of mobile cloudlets for mobile applications in edge computing," *Journal of Supercomputing*, no. 0123456789, 2022, doi: 10.1007/s11227-021-04122-7.

[13] H. Wu et al., "Resolving Multitask Competition for Constrained Resources in Dispersed Computing: A Bilateral Matching Game," *IEEE*

- Internet of Things Journal, vol. 8, no. 23, pp. 16972–16983, 2021, doi: 10.1109/JIOT.2021.3075673.
- [14] R. Singh, S. Armour, A. Khan, M. Sooriyabandara, and G. Oikonomou, “Towards Multi-Criteria Heuristic Optimization for Computational Offloading in Multi-Access Edge Computing,” IEEE International Conference on High Performance Switching and Routing, HPSR, vol. 2021-June, 2021, doi: 10.1109/HPSR52026.2021.9481852.
- [15] S. S. G. Chalapathi, V. Chamola, W. Johal, J. Aryal, and R. Buyya, “Energy and latency aware mobile task assignment for green cloudlets,” Simulation Modelling Practice and Theory, vol. 118, p. 102531, 2022, doi: 10.1016/j.simpat.2022.102531.
- [16] H. Ye, F. Huang, and W. Hao, “On Cost-Aware Heterogeneous Cloudlet Deployment for Mobile Edge Computing,” International Journal of Information Technology and Web Engineering, vol. 17, no. 1, pp. 1–23, 2022, doi: 10.4018/ijitwe.297968.
- [17] C. He, R. Wang, D. Wu, H. Zhang, and Z. Tan, “QoS-aware hybrid cloudlet placement over joint fiber and wireless backhaul access network,” Optical Switching and Networking, vol. 45, no. April, p. 100678, 2022, doi: 10.1016/j.osn.2022.100678.
- [18] J. Blank and K. Deb, “A Running Performance Metric and Termination Criterion for Evaluating Evolutionary Multi- And Many-objective Optimization Algorithms,” 2020 IEEE Congress on Evolutionary Computation, CEC 2020 - Conference Proceedings, pp. 3–10, 2020, doi: 10.1109/CEC48606.2020.9185546.
- [19] H. Song, B. Gu, K. Son, and W. Choi, “Joint Optimization of Edge Computing Server Deployment and User Offloading Associations in Wireless Edge Network via a Genetic Algorithm,” IEEE Transactions on Network Science and Engineering, vol. 9, no. 4, pp. 2535–2548, 2022, doi: 10.1109/TNSE.2022.3165372.
- [20] M. Jia, J. Cao, and W. Liang, “Optimal Cloudlet Placement and User to Cloudlet Allocation in Wireless Metropolitan Area Networks,” IEEE Transactions on Cloud Computing, vol. 5, no. 4, pp. 725–737, 2017, doi: 10.1109/tcc.2015.2449834.
- [21] F. Zeng, Y. Ren, X. Deng, and W. Li, “Cost-effective edge server placement in wireless metropolitan area networks,” Sensors (Switzerland), vol. 19, no. 1, pp. 1–21, 2019, doi: 10.3390/s19010032.
- [22] M. Jia, W. Liang, Z. Xu, M. Huang, and Y. Ma, “QoS-Aware Cloudlet Load Balancing in Wireless Metropolitan Area Networks,” IEEE Transactions on Cloud Computing, vol. 8, no. 2, pp. 623–634, 2020, doi: 10.1109/TCC.2017.2786738.
- [23] N. Hassan, K. L. A. Yau, and C. Wu, “Edge computing in 5G: A review,” IEEE Access, vol. 7, pp. 127276–127289, 2019, doi: 10.1109/ACCESS.2019.2938534.
- [24] S. B. M. Baskaran and G. Raja, “Blind key distribution mechanism to secure wireless metropolitan area network,” CSI Transactions on ICT, vol. 4, no. 2–4, pp. 157–163, 2016, doi: 10.1007/s40012-016-0110-3.
- [25] M. L. Attiah, A. A. M. Isa, Z. Zakaria, M. K. Abdulhameed, M. K. Mohsen, and A. M. Dinar, “Independence and Fairness Analysis of 5G mmWave Operators Utilizing Spectrum Sharing Approach,” Mobile Information Systems, vol. 2019, 2019, doi: 10.1155/2019/4370847.
- [26] X. Xu et al., “An energy-aware computation offloading method for smart edge computing in wireless metropolitan area networks,” Journal of Network and Computer Applications, vol. 133, no. September 2018, pp. 75–85, 2019, doi: 10.1016/j.jnca.2019.02.008.
- [27] S. Wang, Y. Zhao, J. Xu, J. Yuan, and C. H. Hsu, “Edge server placement in mobile edge computing,” Journal of Parallel and Distributed Computing, vol. 127, pp. 160–168, 2019, doi: 10.1016/j.jpdc.2018.06.008.
- [28] F. Guo, B. Tang, and J. Zhang, “Mobile edge server placement based on meta-heuristic algorithm,” Journal of Intelligent and Fuzzy Systems, vol. 40, no. 5, pp. 8883–8897, 2021, doi: 10.3233/JIFS-200933.
- [29] X. Xu et al., “A computation offloading method over big data for IoT-enabled cloud-edge computing,” Future Generation Computer Systems, vol. 95, pp. 522–533, 2019, doi: 10.1016/j.future.2018.12.055.
- [30] B. Li, P. Hou, H. Wu, R. Qian, and H. Ding, “Placement of edge server based on task overhead in mobile edge computing environment,” Transactions on Emerging Telecommunications Technologies, vol. November, pp. 1–19, 2020, doi: 10.1002/ett.4196.
- [31] L. Chen, J. Wu, G. Zhou, and L. Ma, “QUICK: QoS-guaranteed efficient cloudlet placement in wireless metropolitan area networks,” Journal of Supercomputing, vol. 74, no. 8, pp. 4037–4059, 2018, doi: 10.1007/s11227-018-2412-8.
- [32] Z. Qin, F. Xu, Y. Xie, Z. Zhang, and G. Li, “An improved Top-K algorithm for edge servers deployment in smart city,” Transactions on Emerging Telecommunications Technologies, vol. 32, no. 8, pp. 1–20, 2021, doi: 10.1002/ett.4249.
- [33] K. Cao, L. Li, Y. Cui, T. Wei, and S. Hu, “Exploring Placement of Heterogeneous Edge Servers for Response Time Minimization in Mobile Edge-Cloud Computing,” IEEE Transactions on Industrial Informatics, vol. 17, no. 1, pp. 494–503, 2021, doi: 10.1109/TII.2020.2975897.
- [34] Q. Fan and N. Ansari, “On cost aware cloudlet placement for mobile edge computing,” IEEE/CAA Journal of Automatica Sinica, vol. 6, no. 4, pp. 926–937, 2019, doi: 10.1109/JAS.2019.1911564.
- [35] L. Cui et al., “Joint optimization of energy consumption and latency in mobile edge computing for internet of things,” IEEE Internet of Things Journal, vol. 6, no. 3, pp. 4791–4803, 2019, doi: 10.1109/JIOT.2018.2869226.
- [36] X. Li, F. Zeng, G. Fang, Y. Huang, and X. Tao, “Load balancing edge server placement method with QoS requirements in wireless metropolitan area networks,” no. iii, 2021, doi: 10.1049/iet-com.2020.0651.
- [37] L. Zhao, W. Sun, Y. Shi, and J. Liu, “Optimal Placement of Cloudlets for Access Delay Minimization in SDN-Based Internet of Things Networks,” IEEE Internet of Things Journal, vol. 5, no. 2, pp. 1334–1344, 2018, doi: 10.1109/JIOT.2018.2811808.
- [38] S. Mondal, G. Das, and E. Wong, “Efficient cost-optimization frameworks for hybrid cloudlet placement over fiber-wireless networks,” Journal of Optical Communications and Networking, vol. 11, no. 8, pp. 437–451, 2019, doi: 10.1364/JOCN.11.000437.
- [39] K. Peng et al., “An energy- and cost-aware computation offloading method for workflow applications in mobile edge computing,” Eurasip Journal on Wireless Communications and Networking, vol. 2019, no. 1, 2019, doi: 10.1186/s13638-019-1526-x.
- [40] Liqing Liu, Zheng Chang, Xijuan Guo, and T. Ristaniemi, “Multi-objective optimization for computation offloading in mobile-edge computing,” in 2017 IEEE Symposium on Computers and Communications (ISCC), Jul. 2017, vol. 5, no. 1, pp. 832–837, doi: 10.1109/ISCC.2017.8024630.
- [41] Z. Wang, F. Gao, and X. Jin, “Optimal deployment of cloudlets based on cost and latency in Internet of Things networks,” Wireless Networks, vol. 26, no. 8, pp. 6077–6093, 2020, doi: 10.1007/s11276-020-02418-9.
- [42] J. Liu et al., “Reliability-Enhanced Task Offloading in Mobile Edge Computing Environments,” IEEE Internet of Things Journal, vol. PP, no. c, pp. 1–1, 2021, doi: 10.1109/JIOT.2021.3115807.
- [43] F. Guo, B. Tang, L. Kang, and L. Zhang, “Mobile Edge Server Placement Based on Bionic Swarm Intelligent Optimization Algorithm,” in Lecture Notes of the Institute for Computer Sciences, Social- Informatics and Telecommunications Engineering, LNICST, 2021, vol. 350, pp. 95–111, doi: 10.1007/978-3-030-67540-0\_6.
- [44] X. Zhao, Y. Zeng, H. Ding, B. Li, and Z. Yang, “Optimize the placement of edge server between workload balancing and system delay in smart city,” Peer-to-Peer Networking and Applications, vol. 14, no. 6, pp. 3778–3792, Nov. 2021, doi: 10.1007/s12083-021-01208-0.
- [45] H. Wei, H. Luo, Y. Sun, and M. S. Obaidat, “Value-driven Cache Replacement Strategy in Mobile Edge Computing,” 2020 IEEE Global Communications Conference, GLOBECOM 2020 - Proceedings, vol. 2020-Janua, pp. 19–24, 2020, doi: 10.1109/GLOBECOM42002.2020.9348106.
- [46] Z. Liu, J. Zhang, and J. Wu, “Joint Optimization of Server Placement and Content Caching in Mobile Edge Computing Networks,” ACM International Conference Proceeding Series, pp. 149–153, 2019, doi: 10.1145/3375998.3376024.
- [47] T. L. Chin, Y. S. Chen, and K. Y. Lyu, “Queueing Model Based Edge Placement for Work Offloading in Mobile Cloud Networks,” IEEE Access, vol. 8, pp. 47295–47303, 2020, doi: 10.1109/ACCESS.2020.2979479.

- [48] X. Xu, X. Zhang, H. Gao, Y. Xue, L. Qi, and W. Dou, "BeCome: Blockchain-Enabled Computation Offloading for IoT in Mobile Edge Computing," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4187–4195, 2020, doi: 10.1109/TII.2019.2936869.
- [49] C. Zhang and Z. Zheng, "Task migration for mobile edge computing using deep reinforcement learning," *Future Generation Computer Systems*, vol. 96, pp. 111–118, 2019, doi: 10.1016/j.future.2019.01.059.
- [50] L. Loven et al., "Scaling up an Edge Server Deployment," in *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Mar. 2020, pp. 1–7, doi: 10.1109/PerComWorkshops48775.2020.9156204.
- [51] H. Yao, C. Bai, M. Xiong, D. Zeng, and Z. Fu, "Heterogeneous cloudlet deployment and user-cloudlet association toward cost effective fog computing," *Concurrency Computation*, vol. 29, no. 16, pp. 1–9, 2017, doi: 10.1002/cpe.3975.
- [52] J. Zhang, M. Li, X. Zheng, and C. H. Hsu, "A Time-Driven Cloudlet Placement Strategy for Workflow Applications in Wireless Metropolitan Area Networks," *Sensors*, vol. 22, no. 9, pp. 1–19, 2022, doi: 10.3390/s22093422.
- [53] H. Zhang, R. Wang, W. Sun, and H. Zhao, "Mobility Management for Blockchain-based Ultra-dense Edge Computing: A Deep Reinforcement Learning Approach," *IEEE Transactions on Wireless Communications*, vol. 1276, no. c, pp. 1–14, 2021, doi: 10.1109/TWC.2021.3082986.
- [54] Y. Miao, G. Wu, M. Li, A. Ghoneim, and M. Al-rakhami, "Intelligent task prediction and computation offloading based on mobile-edge cloud computing," *Future Generation Computer Systems*, vol. 102, pp. 925–931, 2020, doi: 10.1016/j.future.2019.09.035.
- [55] K. Peng, X. Qian, B. Zhao, K. Zhang, and Y. Liu, "A New Cloudlet Placement Method Based on Affinity Propagation for Cyber-Physical-Social Systems in Wireless Metropolitan Area Networks," *IEEE Access*, vol. 8, pp. 34313–34325, 2020, doi: 10.1109/ACCESS.2020.2974895.
- [56] Z. Xu, W. Liang, W. Xu, M. Jia, and S. Guo, "Efficient Algorithms for Capacitated Cloudlet Placements," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 10, pp. 2866–2880, 2016, doi: 10.1109/TPDS.2015.2510638.
- [57] Z. Xu et al., "An IoT-oriented offloading method with privacy preservation for cloudlet-enabled wireless metropolitan area networks," *Sensors (Switzerland)*, vol. 18, no. 9, pp. 1–18, 2018, doi: 10.3390/s18093030.
- [58] I. Hadžić, Y. Abe, and H. C. Woithe, "Server Placement and Selection for Edge Computing in the ePC," *IEEE Transactions on Services Computing*, vol. 12, no. 5, pp. 671–684, 2019, doi: 10.1109/TSC.2018.2850327.
- [59] S. Mondal, G. Das, and E. Wong, "Cost-optimal cloudlet placement frameworks over fiber-wireless access networks for low-latency applications," *Journal of Network and Computer Applications*, vol. 138, no. March, pp. 27–38, 2019, doi: 10.1016/j.jnca.2019.04.014.
- [60] T. K. Rodrigues, K. Suto, and N. Kato, "Edge Cloud Server Deployment with Transmission Power Control through Machine Learning for 6G Internet of Things," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 4, pp. 2099–2108, 2019, doi: 10.1109/TETC.2019.2963091.
- [61] S. Yang, F. Li, M. Shen, X. Chen, X. Fu, and Y. Wang, "Cloudlet placement and task allocation in mobile edge computing," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5853–5863, 2019, doi: 10.1109/JIOT.2019.2907605.
- [62] S. Bi, L. Huang, and Y. J. A. Zhang, "Joint optimization of service caching placement and computation offloading in mobile edge computing systems," *arXiv*, vol. 19, no. 7, pp. 4947–4963, 2019, doi: 10.1109/TWC.2020.2988386.
- [63] N. Shan, Y. Li, and X. Cui, "A Multilevel Optimization Framework for Computation Offloading in Mobile Edge Computing," *Mathematical Problems in Engineering*, vol. 2020, 2020, doi: 10.1155/2020/4124791.
- [64] B. Li, P. Hou, H. Wu, and F. Hou, "Optimal edge server deployment and allocation strategy in 5G ultra-dense networking environments," vol. 72, 2021, doi: 10.1016/j.pmcj.2020.101312.
- [65] D. Bhatta and L. Mashayekhy, "A Bifactor Approximation Algorithm for Cloudlet Placement in Edge Computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 8, pp. 1787–1798, 2022, doi: 10.1109/TPDS.2021.3126256.
- [66] Y. Li, A. Zhou, X. Ma, and S. Wang, "Profit-Aware Edge Server Placement," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 55–67, Jan. 2022, doi: 10.1109/JIOT.2021.3082898.
- [67] U. Shaukat, E. Ahmed, Z. Anwar, and F. Xia, "Cloudlet deployment in local wireless networks: Motivation, architectures, applications, and open challenges," *Journal of Network and Computer Applications*, vol. 62, no. DECEMBER, pp. 18–40, 2016, doi: 10.1016/j.jnca.2015.11.009.
- [68] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, "Edge computing: A survey," *Future Generation Computer Systems*, vol. 97, pp. 219–235, 2019, doi: 10.1016/j.future.2019.02.050.
- [69] B. Wang, C. Wang, W. Huang, Y. Song, and X. Qin, "A survey and taxonomy on task offloading for edge-cloud computing," *IEEE Access*, vol. 8, pp. 186080–186101, 2020, doi: 10.1109/ACCESS.2020.3029649.
- [70] A. Shakarami, M. Ghobaei-Arani, and A. Shahidinejad, "A survey on the computation offloading approaches in mobile edge computing: A machine learning-based perspective," *Computer Networks*, vol. 182, no. March, 2020, doi: 10.1016/j.comnet.2020.107496.
- [71] V. Pardo-Castello and F. R. Tiant, "Multi-Objective Optimization Using Evolutionary Algorithms: An Introduction," *Journal of the American Medical Association*, vol. 121, no. 16, pp. 1264–1269, 2011, doi: 10.1007/978-0-85729-652-8\_1.
- [72] A. Rahbari, M. M. Nasiri, F. Werner, M. M. Musavi, and F. Jolai, "The vehicle routing and scheduling problem with cross-docking for perishable products under uncertainty: Two robust bi-objective models," *Applied Mathematical Modelling*, vol. 70, pp. 605–625, 2019, doi: 10.1016/j.apm.2019.01.047.
- [73] M. L. Ryerkerk, R. C. Averill, K. Deb, and E. D. Goodman, "Solving metameric variable-length optimization problems using genetic algorithms," *Genetic Programming and Evolvable Machines*, vol. 18, no. 2, pp. 247–277, 2017, doi: 10.1007/s10710-016-9282-8.
- [74] G. R. Zavala, A. J. Nebro, F. Luna, and C. A. Coello Coello, "A survey of multi-objective metaheuristics applied to structural optimization," *Structural and Multidisciplinary Optimization*, vol. 49, no. 4, pp. 537–558, 2014, doi: 10.1007/s00158-013-0996-4.
- [75] A. M. Maia, Y. Ghamri-doudane, D. Vieira, and M. F. De, "An Improved Multi-Objective Genetic Algorithm With Heuristic Initialization for service placement and Load Distribution in Edge Computing," *Computer Networks*, p. 108146, 2021, doi: 10.1016/j.comnet.2021.108146.
- [76] X. Zhu and M. C. Zhou, "Multi-Objective Optimized Cloudlet Deployment and Task Offloading for Mobile Edge Computing," *IEEE Internet of Things Journal*, vol. 4662, no. c, pp. 1–1, 2021, doi: 10.1109/jiot.2021.3073113.
- [77] K. Peng, H. Huang, S. Wan, and V. C. M. Leung, "End-edge-cloud collaborative computation offloading for multiple mobile users in heterogeneous edge-server environment," *Wireless Networks*, vol. 1, 2020, doi: 10.1007/s11276-020-02385-1.
- [78] F. Sufyan and A. Banerjee, "Computation Offloading for Distributed Mobile Edge Computing Network: A Multiobjective Approach," *IEEE Access*, vol. 8, no. Mcc, pp. 149915–149930, 2020, doi: 10.1109/ACCESS.2020.3016046.
- [79] F. Song, H. Xing, S. Luo, D. Zhan, P. Dai, and R. Qu, "A Multiobjective Computation Offloading Algorithm for Mobile-Edge Computing," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8780–8799, 2020, doi: 10.1109/JIOT.2020.2996762.
- [80] F. Luo, S. Zheng, W. Ding, J. Fuentes, and Y. Li, "An Edge Server Placement Method Based on Reinforcement Learning," *Entropy*, vol. 24, no. 3, pp. 1–14, 2022, doi: 10.3390/e24030317.
- [81] Zhilei Ren, He Jiang, Jifeng Xuan, and Zhongxuan Luo, "An Accelerated-Limit-Crossing-Based Multilevel Algorithm for the \$P\$-\$S\$-Median Problem," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 4, pp. 1187–1202, Aug. 2012, doi: 10.1109/TSMCB.2012.2188100.
- [82] Michael R. Garey and D. S. Johnson, *A Guide to the Theory of NP-Completeness*. 1979.
- [83] O. Kariv and S. L. Hakimi, "An Algorithmic Approach to Network Location Problems. I: The p -Centers," *SIAM Journal on Applied*



- Mathematics, vol. 37, no. 3, pp. 513–538, Dec. 1979, doi: 10.1137/0137040.
- [84] A. M. Jubair, R. Hassan, A. H. M. Aman, and H. Sallehudin, “Social class particle swarm optimization for variable-length Wireless Sensor Network Deployment,” *Applied Soft Computing*, vol. 113, p. 107926, Dec. 2021, doi: 10.1016/j.asoc.2021.107926.
- [85] T. Dokeroglu, E. Sevinc, T. Kucukyilmaz, and A. Cosar, “A survey on new generation metaheuristic algorithms,” *Computers & Industrial Engineering*, vol. 137, no. September, p. 106040, Nov. 2019, doi: 10.1016/j.cie.2019.106040.
- [86] M. Ryerkerk, R. Averill, K. Deb, and E. Goodman, “A novel selection mechanism for evolutionary algorithms with metameric variable-length representations,” *Soft Computing*, vol. 24, no. 21, pp. 16439–16452, 2020, doi: 10.1007/s00500-020-04953-1.
- [87] P. Lu, L. Ye, Y. Zhao, B. Dai, M. Pei, and Y. Tang, “Review of metaheuristic algorithms for wind power prediction: Methodologies, applications and challenges,” *Applied Energy*, vol. 301, no. June, p. 117446, Nov. 2021, doi: 10.1016/j.apenergy.2021.117446.
- [88] N. Razmjoo, V. V. Estrela, R. Padilha, and A. C. B. Monteiro, “Metaheuristics and Optimization in Computer and Electrical Engineering,” in *Lecture Notes in Electrical Engineering*, vol. 696, 2021, pp. 25–47.
- [89] H. Li, K. Deb, and Q. Zhang, “Variable-length pareto optimization via decomposition-based evolutionary multiobjective algorithm,” *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 6, pp. 987–999, 2019, doi: 10.1109/TEVC.2019.2898886.
- [90] R. Cheng, C. He, Y. Jin, and X. Yao, “Model-based evolutionary algorithms: a short survey,” *Complex & Intelligent Systems*, vol. 4, no. 4, pp. 283–292, 2018, doi: 10.1007/s40747-018-0080-1.
- [91] R. Lahoz-Beltra, “Quantum Genetic Algorithms for Computer Scientists,” *Computers*, vol. 5, no. 4, p. 24, Oct. 2016, doi: 10.3390/computers5040024.
- [92] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by Simulated Annealing,” *Science*, vol. 220, no. 4598, pp. 671–680, May 1983, doi: 10.1126/science.220.4598.671.
- [93] P. C. Huang, T. L. Chin, and T. Y. Chuang, “Server Placement and Task Allocation for Load Balancing in Edge-Computing Networks,” *IEEE Access*, vol. 9, pp. 138200–138208, 2021, doi: 10.1109/ACCESS.2021.3117870.
- [94] S. K. Kasi et al., “Heuristic Edge Server Placement in Industrial Internet of Things and Cellular Networks,” *IEEE Internet of Things Journal*, vol. 4662, no. c, pp. 1–8, 2020, doi: 10.1109/JIOT.2020.3041805.
- [95] S. Mirjalili and A. Lewis, “The Whale Optimization Algorithm,” *Advances in Engineering Software*, vol. 95, pp. 51–67, May 2016, doi: 10.1016/j.advengsoft.2016.01.008.
- [96] M. Huang, Q. Zhai, Y. Chen, S. Feng, and F. Shu, “Multi-objective whale optimization algorithm for computation offloading optimization in mobile edge computing,” *Sensors*, vol. 21, no. 8, pp. 1–24, 2021, doi: 10.3390/s21082628.
- [97] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of ICNN’95 - International Conference on Neural Networks*, 1995, vol. 4, pp. 1942–1948, doi: 10.1109/ICNN.1995.488968.
- [98] M. Dorigo, V. Maniezzo, and A. Colomi, “Ant system: optimization by a colony of cooperating agents,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 26, no. 1, pp. 29–41, Feb. 1996, doi: 10.1109/3477.484436.
- [99] Y. Dai, Y. Lou, and X. Lu, “A Task Scheduling Algorithm Based on Genetic Algorithm and Ant Colony Optimization Algorithm with Multi-QoS Constraints in Cloud Computing,” in *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, Aug. 2015, vol. 2, no. 1, pp. 428–431, doi: 10.1109/IHMSC.2015.186.
- [100] X. Yang and Suash Deb, “Cuckoo Search via Levy flights,” in *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, 2009, pp. 210–214, doi: 10.1109/NABIC.2009.5393690.
- [101] H. Li, P. G. H. Nichols, S. Han, K. J. Foster, K. Sivasithamparam, and M. J. Barbetti, “Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces,” *Australasian Plant Pathology*, vol. 38, no. 3, p. 284, 1997, doi: 10.1023/A:1008202821328.
- [102] M. Gendreau and J.-Y. Potvin, “Tabu Search,” in *Search Methodologies*, vol. 1–2, Boston, MA: Springer US, 2005, pp. 165–186.
- [103] M. Ryerkerk, R. Averill, K. Deb, and E. Goodman, A survey of evolutionary algorithms using metameric representations, vol. 20, no. 4. Springer US, 2019.
- [104] Z. Qiongbing and D. Lixin, “A new crossover mechanism for genetic algorithms with variable-length chromosomes for path optimization problems,” *Expert Systems with Applications*, vol. 60, pp. 183–189, 2016, doi: 10.1016/j.eswa.2016.04.005.
- [105] B. Wang, Y. Sun, B. Xue, and M. Zhang, “Evolving Deep Convolutional Neural Networks by Variable-Length Particle Swarm Optimization for Image Classification,” *2018 IEEE Congress on Evolutionary Computation, CEC 2018 - Proceedings*, pp. 1–8, 2018, doi: 10.1109/CEC.2018.8477735.
- [106] H. ZainEldin, M. Badawy, M. Elhosseini, H. Arafat, and A. Abraham, “An improved dynamic deployment technique based-on genetic algorithm (IDDT-GA) for maximizing coverage in wireless sensor networks,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 10, pp. 4177–4194, 2020, doi: 10.1007/s12652-020-01698-5.
- [107] V. P. Ha, T. K. Dao, N. Y. Pham, and M. H. Le, “A variable-length chromosome genetic algorithm for time-based sensor network schedule optimization,” *Sensors*, vol. 21, no. 12, pp. 1–25, 2021, doi: 10.3390/s21123990.

# Proof-of-Work for Merkle based Access Tree in Patient Centric Data

B Ravinder Reddy<sup>1</sup>, T Adilakshmi<sup>2</sup>

Research Scholar, Department of CSE, UCE (A), OU<sup>1</sup>

Assistant Professor, Department of CSE, Anurag University, Hyderabad, T.S. India<sup>1</sup>

Professor and Head, Department of CSE, Vasavi College of Engg. (A), Hyderabad. T.S. India<sup>2</sup>

**Abstract**—With the advent of wearable devices and smart health care, the wearable health care technology for obtaining Patient Centric Data (PCD) has gained popularity in recent years. To establish access control over encrypted data in health records, Ciphertext Policy-Attribute Based Encryption (CP-ABE), is used. The most critical element is granting secure access to the generated information. However, with growing complexity of access policy, computational overhead of encryption and decryption process also increases. As a result, ensuring data access control as well as efficiency in PCD collected by wearables is crucial and challenging. This paper proposes and demonstrates a proof-of-work for the Merkle-based access tree using notion of hiding the sensitive access policy attributes.

**Keywords**—Merkle tree; hashing; CP-ABE; access policy; PCD

## I. INTRODUCTION

In order to manage access to important or valuable resources, access control is governed in computer security, according to Di Francesco Maesa et al. [1]. When access requests are made, access control policies are matched to the existing access context to determine subjects' permissions to resources. Many access control methods have been established in research to prevent unauthorized access to vital assets that are more centralized than distributed, such as healthcare records. According to Hong-Ning Dai et al. [2], Healthcare frameworks will inevitably undergo a digital change. The Internet of Medical Things (IoMT) is critical, but inherent security and privacy concerns limit mainstream use. According to Vincent Hu et al., [3], Blockchain has provided new insights into the security of clinical information, with the highlights of distributed, trustless data storage, peer-to-peer transmission, and encryption algorithms, and it may overcome the inconsistency between information sharing and privacy with suitable security measures. Using EHR raises issues regarding the confidentiality, protection, and dependability of medical information, according to Liang Huang et al. [4]. According to Buket Yüksel et al., [5] the combination of Blockchain with cloud has drawn a lot of attention recently for ensuring effectiveness, transparency, security, and providing better cloud administrations. To maximize the capability of the Blockchain-cloud mix, a full understanding of the present initiatives in this field is essential. Since healthcare applications are a good example of dispersed environments and depend on servers for storage and processing, CP-ABE is one of the most promising security mechanisms in this area. One-to-many communication [6] is supported by CP-ABE, and the owner defines the access policy, such that only

approved decryptor can access the data. An encrypted storage system is a use case for CP-ABE. One encryption key cannot encrypt many sets of data, preventing the implementation of fine-grained access control. The CP-ABE approach may provide access control without increasing the amount of keys by indicating the set of properties of the decryptor, such as affiliation. Enhanced schemes of ABE include broadcast encryption, multi-authority provisioning, and anonymous user or decryptor. Additionally, the number of attributes and pairing computations are inversely related. On the other hand, Merkle trees are designed to authenticate messages with a distinctive signature while also enabling an intended validator to confirm the authenticity of one message without exposing the validity of others.

## A. Research Problem

One of the main challenges in managing access to sensitive patient information is the attribute disclosure problem. Traditional access control methods, such as role-based access control (RBAC) and access control lists (ACLs), are not well-suited to the attribute disclosure problem, as they do not provide fine-grained control over access to sensitive patient information based on the user's attributes. CPABE is one of the few models that can provide this type of access control, but the attribute disclosure problem still persists as the access policy is exposed. To overcome this issue, a novel approach of attribute hiding through Merkle tree is proposed for PCD while preserving privacy. The Proof-of-work ensures that the decryption is performed by an authorized user and not an attacker.

## B. Research Questions

- 1) What are the main challenges in managing access to sensitive patient information?
- 2) How does the proposed Merkle tree based access structure for sensitive attributes of PCD in CPABE avoid unauthorized access and preserve privacy?
- 3) How can cryptographic techniques be used to ensure that the Merkle tree-based CPABE system is verifiable?
- 4) What are the benefits of a verifiable Merkle tree-based CPABE system in terms of in security and privacy?

## C. Research Objectives

The research objective of this study is to address the main challenges in managing access to sensitive patient information by proposing a Merkle tree based access structure for sensitive attributes of PCD in CPABE, and examining how this

structure can be used to avoid unauthorized access and preserve privacy. Additionally, the objective is to investigate the use of cryptographic techniques to ensure that the Merkle tree-based CPABE system is verifiable and to explore the benefits of such a system in terms of security and privacy.

#### D. Research Significance

The Proof-of-work proposed in the article provides a secure method for managing sensitive attributes in a patient-centric approach. It enhances security and integrity for both users and providers in healthcare associations. The proposed scheme uses a monotonic access structure for policy transmission, and is resistant to chosen-ciphertext attacks (CCA) even when an attacker acquires multiple private keys. Additionally, it can handle large volumes of data and provide quick integrity proof. This model contributes new knowledge in the domain of secure access control for sensitive patient information.

## II. LITERATURE REVIEW

By 2025, the worldwide IoT healthcare industry is anticipated to grow to 534.3 billion US Dollars. Carminati B, et al., [7] wearable health care is a popular and convenient technology that enables users to access medical services. Wearable medical data needs rapid and accurate information sharing from any place for better healthcare choices. However, the security and confidentiality of such patient information will become a major issue while sharing. When unauthorized access to information is unavoidable in medical settings, the security of patients' data is crucial. Siti Dhalila et al., [8] to overcome these challenges, new systems must be created with protected electronic health data, efficient storage, and a properly verified retrieval mechanism. Clauson et al., [9] aside from the patient-provider interaction and access-sharing, such a provision must also protect the patient's privacy with a greater focus on patient-centric health data management. Shan Jiang et al., [10] presented an access control mechanism based on smart contracts for access verification to get EHRs sharing by integrating data dumping and sharing processes for providing e-medical services through cloud and Blockchain. The medical information obtained by IoT devices is forwarded to nearest edge workers for information management while providing security. The data is then shared via Blockchain exchange.

According to C. Nguyen et al., [11], a framework-proposed fast CP-ABE model can assign costly computing jobs to semi-trusted third parties while keeping consistent number of simple calculations. To validate the validity of the decryption output, a Boneh-Lynn-Shacham signature model is utilized. Wang et al., [12] the approach creates a privacy schema for clinical data that is based on Blockchain and the cloud. It can perform safe insurance and clinical information integrity verification, as well as handle computation, information sharing, and security challenges. Wang et al., [13] offer a patient-driven PCD sharing system to preserve patients' privacy and provide them control over their PCDs. Prior to reevaluating, all PCDs in this structure are encrypted with multi-authority and ABE, addressing the key facilitation issue and achieving fine-grained access control. Additionally, anonymous authentication between the cloud and the user is

advised to preserve integrity while concealing the patient during validation. Leyou Zhang et al., [14] demonstrate a successful character identity-based distributed decryption technique for a medical records sharing framework. It allows them to share their data with different people without having to recreate the decryption of their private key. ABE was proposed [15] in order for the user to decode; there must be at least  $n$  attributes that match between the ciphertext and the user. V Goyal et al., [16] design extended the expressiveness of access structure by tying a ciphertext to a set of attributes, which they dubbed KP-ABE. Bethencourt et al., [17] introduced CP-ABE, which employs attributes to reflect the credentials of a user. Healthcare data that is sent in real-time or kept on a third-party cloud server is susceptible to a number of threats. A patient's life might be severely affected by the unauthorized access to sensitive medical information. Communication between multiple devices is one of the criteria for an IoT healthcare system. CP-ABE is a method that has promise for providing role-based access control in this situation. A type of public key cryptosystem called CP-ABE uses a collection of attributes to establish the user roles. In this case, the decryption key is linked to the receiver's attribute set and the access policy is contained with ciphertext. If and only if, the recipient's attribute set complies with the access structure/policy used for encryption, a successful decryption will take place. The data owners established the access policy. According to Hui Cui et al., [18] the decryption keys were only a collection of attributes with no tree structure. In a CP-ABE system, an access structure is included in the ciphertext, which may leak sensitive information about the underlying plaintext and the privileged receivers since anybody who views the ciphertext may learn the privileged recipients' attributes from the associated access structure.

Patients can encrypt data in a variety of methods to tighten access control in healthcare, such that anybody who reads the encrypted text can only comprehend the patient's publicly known qualities and the sensitive information stays hidden. This is according to Vijayan, V et al., [19]. Use of CP-ABE to encrypt the whole access policy with attributes or only the portion of the policy that has to be concealed is one or more solutions. However, the system cannot determine if the end user has the necessary authorization to access it because the ciphertext can only be retrieved by authorized end users, according to Nishide, T et al., [20]. This method's inability to outsource the decryption costs is its second drawback. The system will not be protected by CP-ABE encryption if the policies are wholly or partially obscured. Data must be patient-centric in a healthcare setting since it may be shared across several domains. This may be accomplished by hiding the access policy property using CP-ABE's different degrees of hashing, which supports fine-grained access control. It must also be done efficiently and securely to confirm the existence of the attributes used in the formulation of the access policy without disclosing the content at the time of decryption.

## III. PRELIMINARIES

The access tree is a list of all stakeholders who have access. The proposed approach's major purpose is to prove the existence of attributes in the access structure without revealing the sensitive attributes to the decryptor.

A. System Model

CP-ABE, according to Wang, Shulan, et al. [12], can enable privacy preserving and safe data sharing in public environment. The policy is embedded in the ciphertext, and each user's private key is based on attributes set and successful decryption can occur, iff the ciphertext or attributes satisfy the access key. Fig. 1 shows the Architecture for CP-ABE policy.

B. Access Tree for PCD

The CP-ABE access policy with attributes like hospital,

physician, Insurance company and department is defined for accessing PCD as in figure 2 is (“Physician” AND “Hospital X”) OR (“Insurance company X” AND “Insurance Department”). The proposed algorithm has five phases:

Phase 1: Initial Setup phase -  $S_c(\lambda, S) \rightarrow (PU_K, MS_K)$

The  $K_{GC}$  generates  $PU_K$  and  $MS_K$ , respectively, given by a  $\lambda$  and  $S$ .

Phase 2: Key Generation phase -  $K_G(PU_K, MS_K, S) \rightarrow S_K$

For  $M$ .  $PU_K$ ,  $MS_K$  and  $S$ , the  $K_{GC}$  generates  $S_K$ .

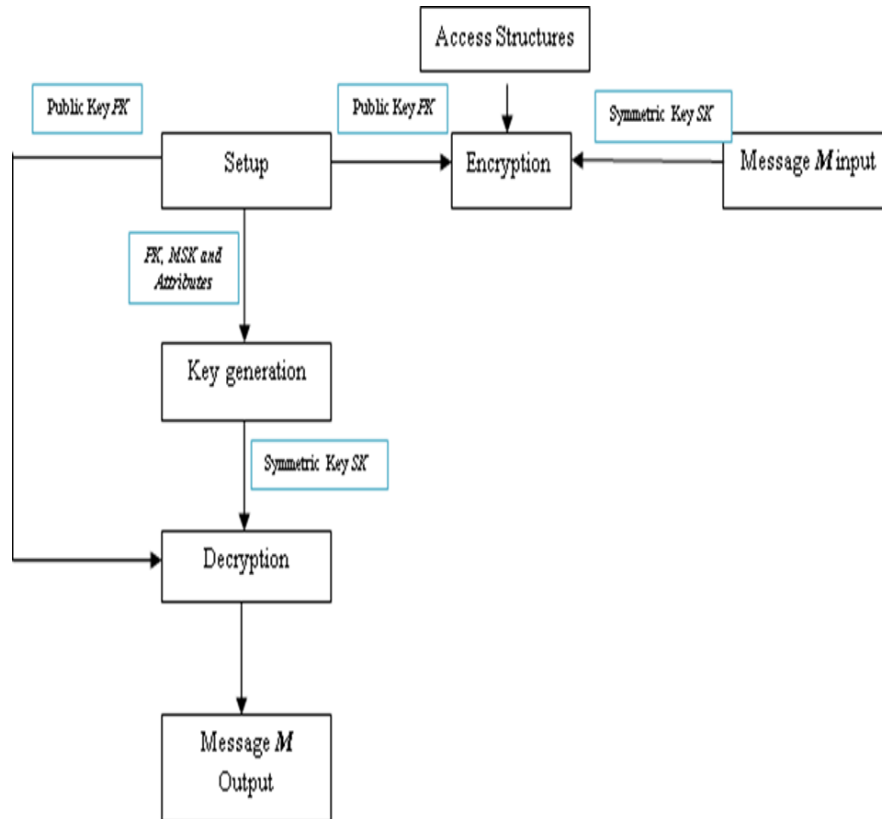


Fig. 1. Traditional CP-ABE architecture.

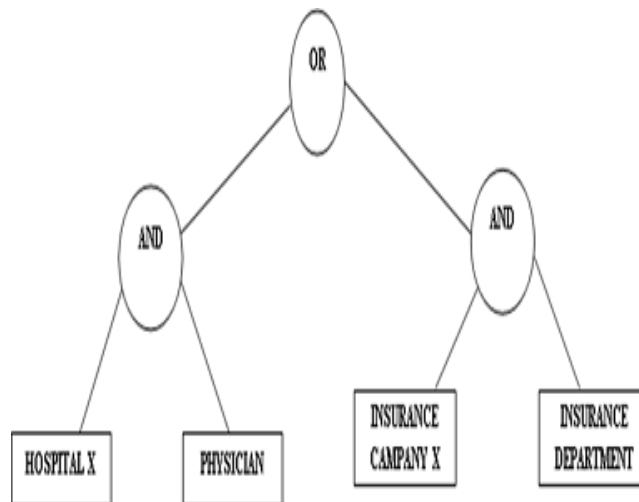


Fig. 2. The access tree generated for PCD.

Phase 3: Access Tree Construction phase

Given a collection of messages,  $M = \{m_1, m_2, \dots, m_n\}$  creates a Merkle tree whose interior nodes include the concatenation of the hash values corresponding to its leaf nodes and whose leaves include the hash value of each message  $m$  in  $M$ . In the Fig. 3 below, shows the Merkle tree construction through attributes set,  $A = \{a_1, a_2, a_3, a_4\}$ .  $H_R$  represents the root hash value also called Merkle Root.

Phase 4: Encryption phase -  $E(PU_K, S_T, NS_T, M) \rightarrow C_T$

The sender outputs the CT with input PUK, trees ST, NST, and M, where ST is a sensitive Merkle tree and

NST is a non-sensitive public tree.

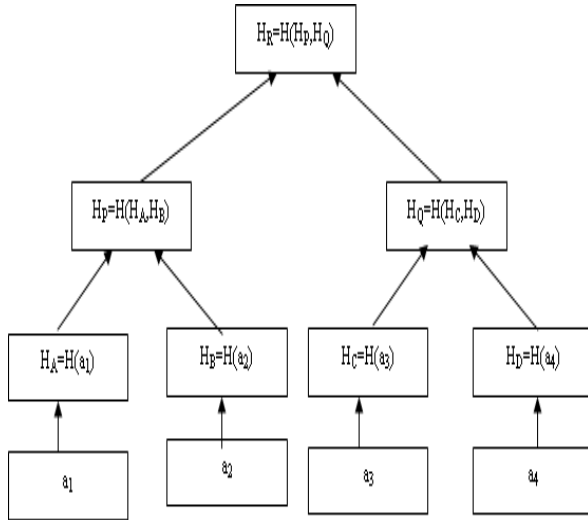


Fig. 3. The sensitive attribute Merkle tree  $S_T$ .

Phase 5: Decryption phase -  $D(C_T, S_K) \rightarrow M$

The decryptor then decrypts  $C_T$  with  $S_K$  and outputs message  $M$ . By moving up the tree  $T_S$  from a specific transaction node to root node, Merkle proofs can demonstrate that a specific attribute is present during decryption. The decryptor examines the public tree  $T_{NS}$  and discovers the Merkle proof for  $S_T$  to verify the integrity of hidden attributes.

Consider  $a_3$  from the attribute set  $A$  and  $H_C$  the hash of  $a_3$  to be verified, represented by bolded rectangle. Given  $H_P$  from other sub tree and  $H_D$ , the adjacent node hash value, represented by dashed rectangles. The  $H_R$  value can be calculated. If  $H_R = H_R$ , the hash root value of the tree when generated as in Fig. 4, the attributes are considered to be secure.

C. Data Set

The tree was designed for use in experiments using the IBM-Clinical Hub [21] data set, which was stated in [23]. This data set contains clinical data elements that, when paired with patient identification data, allow for the storage, production, and access of longitudinal patient records. The IBM-Clinical Hub is a more comprehensive workbench paradigm that offers access to consuming systems and processes as well as patient and clinical data. An ongoing stream of HL7 events and messages is used to construct the longitudinal patient record.

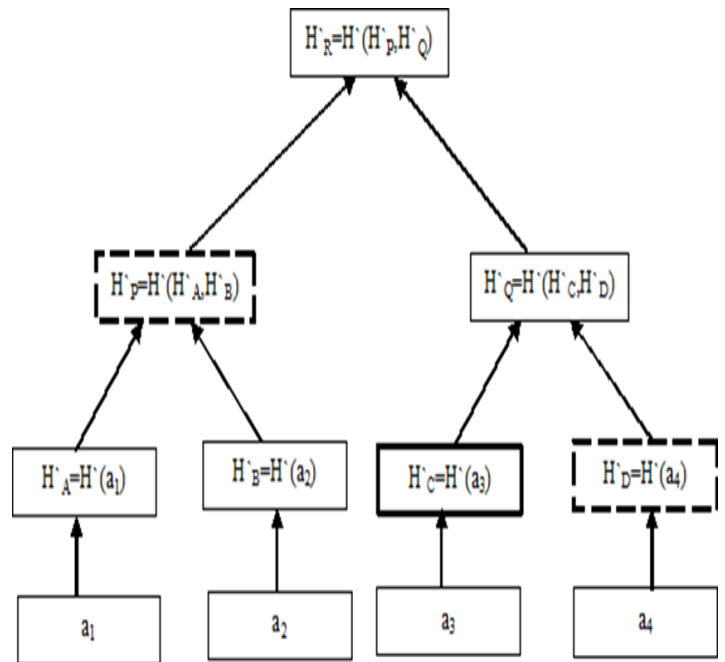


Fig. 4. The proposed sensitive attributes Merkle tree with verifiable nodes.

IV. PROPOSED VERIFIABLE MERKLE PROOF MODEL FOR SENSITIVE ATTRIBUTES

Merkle proof is a type of selective disclosure and is a useful strategy for protecting users' privacy by hiding unwanted sections of certificates or other documents and only offering partial disclosure for evidence. A Merkle proof validates specific transactions represented by a Merkle hash root's leaf or branch hash. Table II represents the list of PCD sensitive attributes for which the tree is generated with the root,  $r$ . The decryptor in the proposed Merkle tree based CP-ABE sensitive attribute access structure scheme will verify the integrity of the non-public attributes by selecting one attribute at random and by giving its hash value as input for the proof. In turn the decryptor will be given few other branch nodes hash values at random. Based on these multiple available hash values, the decryptor constructs the tree resulting to root hash. If the final hash value matches Merkle tree root, it means the attribute exists and the transaction can be processed.

Algorithm for Proposed Model

Step 1:

Consider the access tree [20] ST with root R. Hence,  $ST == TR$ . Where SsT represents the subtree of sensitive attributes tree ST rooted at the node ST.

Step 2:

If  $A(\beta) \in SsT$ , Then denote [15] it as  $SsT(\beta) = 1$ .

Step 3:

Compute  $SsT(\beta)$  recursively as follows.

if(ST  $\neq$  leaf)

Evaluate  $SsT(\beta) = 1 \forall$  children ST' of node ST.

$SsT(\beta) == 1$

iff  $\geq k_s$  with children 1.

if(ST == leaf),

Then  $SsT(\beta) = 1$

iff  $A(ST) \in \beta$ .

V. RESULTS AND DISCUSSION

The proposed Merkle tree based sensitive attribute access structure in CPABE is implemented using the pymerkle package[22], and its performance is evaluated by measuring the time taken to generate the Merkle tree as in Table I and Merkle proof for varying numbers of sensitive attributes using the SHA-256 algorithm. As the number of sensitive attributes increases, the size of the tree and its height also increases, resulting in longer generation times. Additionally, the results show that the Merkle proof computational time increases with the number of sensitive attributes. However, the time taken to generate the tree and Merkle proof decreases as the key size for the SHA-256 algorithm increases. These results demonstrate that the proposed scheme as in Figure 5 is efficient in terms of computational time and can be effectively implemented in real-world scenarios.

A. Comparison with other Frameworks

The Table II represents the comparison of the proposed model with other schemes, covering CPU load, data integrity, proof-of-work, efficiency and privacy. By creating a Merkle tree proof, guaranteeing integrity via a hash and providing effective data access control, the proposed CP-ABE scheme is capable to implement hidden sensitive attribute access policies with less computation load

B. Security Analysis

1) Chosen-ciphertext attack: The proposed Merkle tree model is secure against chosen-ciphertext attack is through the collision-resistant scheme. This means that it is computationally infeasible for an attacker to find two different sets of data blocks that have the same Merkle root.

Let  $H(x)$  be a cryptographic hash function that takes in input  $x$  and produces a fixed-length output. We can construct a Merkle tree with  $n$  leaf nodes, where each leaf node  $i$  is a hash of a data block  $D_i$ , represented as  $H(D_i)$ . The non-leaf nodes in the tree are the hash of their children nodes, represented as  $H(L_i, R_i)$  where  $L_i$  and  $R_i$  are the left and right child nodes respectively.

The Merkle root of the tree, represented as  $R$ , is the hash of the root node and it is computed as  $R = H(H(L_1, R_1), H(L_2, R_2) \dots H(L_{n/2}, R_{n/2}))$ . To show that the tree is collision-resistant, assume that there exists an attacker who can find two different sets of data blocks,  $D_1$  and  $D_2$ , such that  $H(D_1) = H(D_2)$  for all  $i$ . Therefore, the attacker can

construct two different Merkle trees,  $T_1$  and  $T_2$ , with the same Merkle root.

However, since  $H(x)$  is a collision-resistant hash function, it is computationally infeasible for the attacker to find two different inputs  $x_1$  and  $x_2$  such that  $H(x_1) = H(x_2)$ . Therefore, it is also computationally infeasible for the attacker to find two different sets of data blocks that have the same Merkle root, and the Merkle tree structure provides a secure method for verifying the integrity of the sensitive attributes in the access structure. Additionally, the use of a one-way hash function in the construction of the Merkle tree ensures that it is computationally infeasible for an attacker to obtain the original sensitive attributes from the hash values stored in the tree, providing further security against chosen-ciphertext attacks.

TABLE I. TREE GENERATION TIME VS PROOF COMPUTATIONAL TIME

S.No.	Number of Sensitive Attributes	Size/ Total nodes	Tree Height	Merkle Tree generation time (in seconds)*10 <sup>-2</sup>	Merkle Proof computational time (in seconds) *10 <sup>-2</sup>
1.	24	47	5	97.6	23.4
2.	90	179	7	33.0	30.1

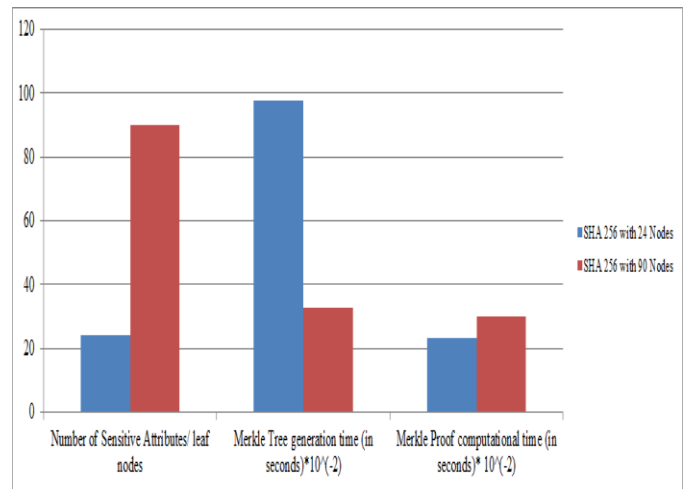


Fig. 5. The sensitive attributes vs tree generation and proof.

TABLE II. COMPARISON OF PROPOSED MODEL WITH OTHER FRAMEWORKS

S.No.	Model	CPU Load	Integrity	Efficiency	Proof of work	Privacy
1.	Siti Dhalila et al., [6]	Moderate	Yes	Improved	No	Through fully hiding Full Policy/ file hiding through one-way Hashing
2.	John Bethencourt et al.,[15]	Moderate	No	Nominal	No	The model applicable to large universe of attributes. Policy and Attribute are exposed to unauthorized
3.	Proposed model	Less	Yes	High	Yes	Partial sensitive attribute hiding

2) *Replay attack*: In order to protect against replay attacks, one approach is to include a timestamp or nonce in the data block that is hashed to create the leaf node of the Merkle tree. When a user requests access to a sensitive attribute, the system will re-compute the hash value of the leaf node using the current timestamp or nonce. This ensures that each request is unique and cannot be replayed at a later time.

For example, consider a user requests access to a sensitive attribute at time  $t_1$ , the system will compute the hash of the data block with the timestamp  $t_1$ . If the same user tries to access the same attribute again at time  $t_2$ , the system will compute the hash of the data block with the timestamp  $t_2$ , which will be different than the previous hash computed at time  $t_1$ , hence it will prevent replay attack as the hash values are different.

## VI. CONCLUSION

The Merkle tree based access structure in CPABE addresses the attribute disclosure problem in managing access to sensitive patient information by providing fine-grained control and preserving privacy through attribute hiding. The proposed model, a proof-of-work algorithm, enhances security and integrity for both users and providers in healthcare organizations by using a monotonic access structure for policy transmission and being resistant to chosen-ciphertext attacks and replay attacks. It can also handle large volumes of data and provide quick integrity proof. While the model addresses limitations of existing solutions and provides a secure solution for healthcare organizations, further research is needed to explore its potential in addressing other security challenges and improving its efficiency. Overall, this study makes a significant contribution to the field of secure access control for sensitive patient information.

## REFERENCES

- [1] Maesa, Damiano di Francesco, Paolo Mori and Laura Ricci. "Blockchain Based Access Control." IFIP International Conference on Distributed Applications and Interoperable Systems (2017).
- [2] H. -N. Dai, M. Imran and N. Haider, "Blockchain-Enabled Internet of Medical Things to Combat COVID-19," in IEEE Internet of Things Magazine, vol. 3, no. 3, pp. 52-57, September 2020.
- [3] Hu, Vincent, Tim Grance, David Ferraiolo, and D. Kuhn. "An Access Control Scheme for Big Data Processing". In Proceedings of the 10th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, ICST, 2014. [4] Huang, Liang, Hyung-Hyo Lee, and Hongju Cheng. "A Medical Data Privacy Protection Scheme Based on Blockchain and Cloud Computing". Wirel. Commun. Mob. Comput. January 2020.
- [4] Yüksel, B., Kıpçü, A., & Özkasap, Ö. "Research issues for privacy and security of electronic health services". Future Gener. Comput. Syst., 68:1–13, 2017.
- [5] Barbara Carminati. Merkle Trees. Encyclopedia of Database Systems. 2009.
- [6] Siti Dhalila Mohd Satar, Mohamad Afendee Mohamed, Masnida Hussin, Zurina Mohd Hanapi and Siti Dhalila Mohd Satar, "Cloud-based Secure Healthcare Framework by using Enhanced Ciphertext Policy Attribute-Based Encryption Scheme" International Journal of Advanced Computer Science and Applications(IJACSA), 12(6), 2021.
- [7] Engelhardt, Mark A.. "Hitching Healthcare to the Chain: An Introduction to Blockchain Technology in the Healthcare Sector." Technology Innovation Management Review 7 (2017): 22-34.
- [8] S. Jiang, J. Cao, H. Wu, Y. Yang, M. Ma and J. He, "BlocHIE: A BLOCkchain-Based Platform for Healthcare Information Exchange," 2018 IEEE International Conference on Smart Computing (SMARTCOMP), Taormina, Italy, 2018, pp. 49-56.
- [9] Dinh C. Nguyen, Pubudu N. Pathirana, Ming Ding, Aruna Seneviratne. "A cooperative architecture of data offloading and sharing for smart healthcare with blockchain". 2021. IEEE International Conference on Blockchain and Cryptocurrency, ICBC 2021.
- [10] Shulan Wang, Haiyan Wang, Jianqiang Li, Huihui Wang, Junaid Chaudhry, Mamoun Alazab, Houbing Song. "A Fast CP-ABE System for Cyber-Physical Security and Privacy in Mobile Healthcare Network. IEEE Transactions on Industry Applications". 564. 2020.
- [11] Chang Ji Wang, Xi Lei Xu, Dong Yuan Shi, Wen Long Lin. "An efficient cloud-based personal health records system using attribute-based encryption and anonymous multi-receiver identity-based encryption". Proceedings - 2014 9th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing.
- [12] Leyou Zhang, Yadi Ye, Yi Mu. "Multiauthority Access Control with Anonymous Authentication for Personal Health Record". IEEE Internet of Things Journal. 81. 2021.
- [13] Amit Sahai, Brent Waters. "Fuzzy identity-based encryption". Lecture Notes in Computer Science. 3494. 2005.
- [14] Vipul Goyal, Omkant Pandey, Amit Sahai, Brent Waters. "Attribute-based encryption for fine-grained access control of encrypted data". Proceedings of the ACM Conference on Computer and Communications Security. 2006.
- [15] John Bethencourt, Amit Sahai, Brent Waters. "Ciphertext-policy attribute-based encryption". Proceedings - IEEE Symposium on Security and Privacy. 2007.
- [16] Hui Cui, Robert H. Deng Junzuo Lai, Xun Yi, Surya Nepal. "An efficient and expressive ciphertext-policy attribute-based encryption scheme with partially hidden access structures, revisited". Computer Networks. 133. 2018.
- [17] Vini Vijayan, James Connolly, Joan Condell, Nigel McKelvey, Philip Gardiner. "Review of wearable devices and data collection considerations for connected health". Sensors 2116. 2021.
- [18] Takashi Nishide, Kazuki Yoneyama, Kazuo Ohta. "Attribute-based encryption with partially hidden encryptor-specified access structures". Lecture Notes in Computer Science. LNCS. 2008.
- [19] "Longitudinal Patient Records Artifacts." Longitudinal Patient Records Artifacts, www.ibm.com, 12 Apr. 2021.
- [20] Nishant Doshi, Devesh Jinwala. "Constant ciphertext length in multi-authority ciphertext policy attribute based encryption". 2011 2nd International Conference on Computer and Communication Technology, ICCCT-2011.
- [21] Nurmamat Helil, Kaysar Rahman. "CP-ABE access control scheme for sensitive data set constraint with hidden access policy and constraint policy". Security and Communication Networks. 2017.
- [22] <https://pymerkle.readthedocs.io/en/latest/index.html#>
- [23] B. Ravinder Reddy, T. Adilakshmi, "Merkle Tree-based Access Structure for Sensitive Attributes in Patient-Centric Data," International Journal of Engineering Trends and Technology, vol. 70, no. 6, pp. 106-113, 2022.

APPENDIX A:

TABLE III. SYMBOL TABLE

S.No.	Symbol	Description
1.	ABE	Attribute Based Encryption
2.	$\{a_1, a_2, a_3 \dots a_n\}$	Attributes
3.	A,B,C,S	Attribute Set
4.	$S_T$	Sensitive attribute Access Tree
5.	$NS_T$	Non-Sensitive attribute Access
6.	$Ss_T$	Sub Tree
7.	T	Threshold value
8.	$K_s$	Child node
9.	$num_s$	Number of children of s
10.	H	Hash function
11.	$R_H$	Root Hash
12.	$\lambda, \beta$	Security Variable
13.	$MS_K$	Master Key
15.	$S_K$	Secret Key
16.	$K_G$	Key generation
17.	$K_{GC}$	Key Generation Centre
18.	PCD	Patient-Centric Data
19.	M	Input Message
20.	HL7	Health Level 7
21.	$C_T$	Cipher Text
22.	E	Encryption
23.	D	Decryption



# Water Tank Wudhu and Monitoring System Design using Arduino and Telegram

Ritzkal<sup>1\*</sup>, Yuggo Afrianto<sup>2</sup>, Indra Riawan<sup>3</sup>, Fitrah Satrya Fajar Kusumah<sup>4</sup>, Dwi Remawati<sup>5</sup>

Universitas Ibn Khaldun, Bogor, Indonesia<sup>1, 2, 3, 4</sup>

STMIK Sinar, Nusantara, Surakarta<sup>5</sup>

**Abstract**—Manual water faucets, which are commonly used in mosques and homes, cannot control water use, resulting in a variety of issues, including water waste when the user forgets to close the water faucet, resulting in water continuously coming out. In addition to filling the water tank, which is also an important factor in saving water, the water reserve in the tank must be properly controlled so that its availability is maintained. Based on the existing problems, a water faucet system was made for abluion and monitoring water tanks using Arduino and Telegram. An automatic abluion water faucet system that can drain water automatically with an ultrasonic sensor as a body movement reader and a solenoid valve as a substitute for a faucet. The water pump can help fill the water tank automatically and know how much water is in the tank using an ultrasonic sensor; liquid crystal display and Telegram as recipients of text messages from the results of the condition of water faucets, water pumps, and water levels.

**Keywords**—Arduino; solenoid valves; ultrasonic sensor; pump water

## I. INTRODUCTION

Water is one of the things that is very influential in human life and is needed in everyday life [1]+ because water is a source of life for humans in this world [2]. In human life, clean water is one of the natural resources used by humans for consumption or in carrying out their daily activities and the existence of water sources should be maintained [3][14]. One of them is the abluion used by Muslims to worship before worship, which requires cleanliness and sanctity [4]. One of the abluion activities is to use clean water that flows over certain limbs, namely the face, hands, head, and feet [5].

In places of worship such as mosques, there are rooms used for abluion, and the faucets used are mostly still manually operated to regulate the size of the water output [6]. To adjust the size and volume of water output, the user must turn the faucet lever manually, and the faucet is easily damaged because the faucet lever is often rotated. so that it becomes one of the factors contributing to excessive water expenditure or wasteful water use [7]. Places of worship, such as mosques, must have tanks or water containers to store clean water [8]. The water reservoir is not only a storage area but also maintains the smoothness and availability of water for the needs of the mosque [9][15]. Filling water and monitoring water tanks in mosques is still done manually, so it often causes waste due to negligence in turning off the water pump when the tank is full [10].

From the problems above, it can be formulated how to regulate the expenditure of water used during abluion so that there is no waste of spending water and the use of water can be controlled, and how to find out the water level in the water tank [11]. For this reason, an open-source electronic platform is needed that is easy to use, namely Arduino. Arduino technology can assist in making things automatic, so they no longer need to be controlled [12]. One way to overcome this problem is to build an automatic abluion water faucet as a regulator of water expenditure by using an ultrasonic sensor as a distance reader and a solenoid valve as a substitute for a manual faucet. In addition, the controller built using ultrasonic sensors can also work in real time, so it is suitable for controlling abluion water in mosques. The use of the microcontroller on Arduino, which is designed by adding the HC-SR04 ultrasonic sensor, solenoid valve, and water pump, is made into a system that works under the control of the user interface and can adjust the opening and closing of water faucets automatically [13] and aims to control the expenditure of water so that it is not excessive and to make it easier for mosque guards to know the contents of the tank.

## II. RESEARCH METHODS

The method in this study is used to carry out a planning stage or describe a workflow in compiling an idea with many stages. The stages used in this research method are as follows:

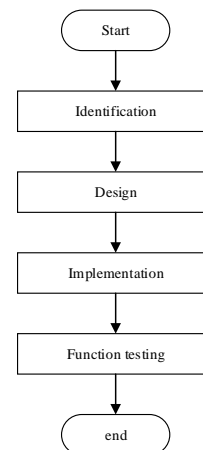


Fig. 1. Research stages

### A. Identification

At this early stage, identification of the requirements that will be needed to design a system is carried out. At this initial stage what is done is to analyze why this research was carried

out. This identification stage is divided into two things, namely needs analysis and work method analysis.

- 1) Identification of needs
- 2) Identify How It Works

### B. Design

The design of this study contains the development of the system work analysis stages, which are changed to block diagrams so that researchers can understand the flow and function of the design that will be made. The following are the stages of this research design. The design of the hardware used in the research is carried out; the design is made in the form of a block diagram to illustrate interconnected diagrams.

### C. Implementation

The implementation phase implements everything that has been well designed, such as hardware design. The hardware implementation stage includes the implementation of connecting modules and the implementation of Telegram.

### D. Function Testing

At this stage, various kinds of tests will be carried out that have been implemented in the previous stage. This stage is carried out by:

- 1) HC-SR04 Ultrasonic Sensor Function Testing
- 2) Relay Function Testing
- 3) Solenoid Valve Function Testing
- 4) Waterpump Function Testing
- 5) Ethernet Shield Testing
- 6) Telegram function testing

## III. RESULT

At this point, the findings of this study, titled "Design of the Ablution Water Faucet System and Monitoring the Water Tank Using Arduino and Telegram," will be discussed.

### A. Identification

Based on the existing problems, it can be concluded from several system requirements regarding the implementation of the design of the abluion water faucet system and monitoring the water tank using Arduino and Telegram, as previously explained, that several needs were needed, including

1) *Identification of needs:* At the stage of identifying the needs that will be carried out, there are several hardware devices to support the implementation of research on the design of the abluion water faucet system and monitoring the water tank using Arduino and telegram, as shown in the following table:

TABLE I. IDENTIFICATION OF REQUIREMENTS (HARDWARE)

No.	Identification of requirements (hardware)
1	Arduino Uno
2	Solenoid Valves
3	Ultrasonic Sensor HC-SR04
4	Water Pump
5	5v Relays
6	Liquid Crystal Display
7	Ethernet Shield
8	Jumper Cables
9	Telegram

In Table I, No. 1, Arduino Uno is a system controller and an electronic prototyping platform that is open-source hardware based on hardware and software that is flexible and easy to use [13]. In Table I, No. 2, Solenoid Valve is a valve that is driven by electrical energy which has a coil as its driving force. In addition, the solenoid valve also has a fast opening and closing response [16]. At No. 3, "The HC-SR04 type ultrasonic sensor is a device used to measure the distance from an object and as an object reader" [17]. In Table I, No. 4, the Water Pump is a water pump motor used for aquariums, fish ponds, hydroponics, robotics, or projects making microcontroller-based applications [18]. In Table I, No. 5, Relays are on-and-off switches that are operated electrically and are electromechanical components [19]. In Table I, No. 6, the Liquid Crystal Display can function to display a sensor result value, display text, or display a menu on the microcontroller [20]. In Table I, No. 7, the Ethernet Shield is a module that is used to connect Arduino to the internet [21]. In Table I, No. 8, Jumper Cables are used to connect all components in the circuit on the breadboard [22]. In Table I, No. 9, Telegram is a cloud-based application, which makes it easier for users to access a Telegram account from different devices and simultaneously [23]. Telegram Messenger is a cross-platform messaging application that allows us to exchange messages without SMS fees because Telegram uses the same internet data package for email, web browsing, and so on [24].

2) *Identify how it works:* In the process of working assistance, it will be explained how the working system works. The following image will explain the analysis of how this system works:

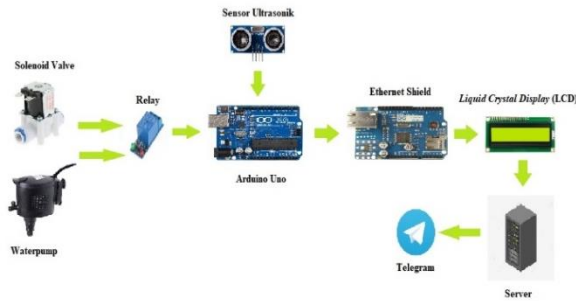


Fig. 2. Working notes

In Fig. 2, explains how the system works, we need to know whether the system we have created is in line with expectations or not. Ultrasonic sensors will be used as input in this research, and data from ultrasonic sensors will be sent to Arduino upon receipt. The relay functions as a limiter on/off the water pump and solenoid valve, while the solenoid valve controls the opening and closing of the faucet. Arduino will process the data and send data output to the LCD in the form of text and numbers. Next, the system will pass the information to telegram via the Ethernet shield network, and then send the information to the server, where the information will be forwarded to a pre-arranged telegram bot, which will automatically route the phone via the telegram messenger.

**B. Design**

The hardware design stages used in the creation of a system are carried out during this design stage. The design is made in the form of a block diagram to illustrate a connected diagram. The following are several design stages in this study:

1) *Hardware design:* Based on the functional block diagram contained in Fig. 3, the system is divided into several parts. Sensors as inputs, Arduino as receivers and senders, relays as on/off switches that act as power regulators that enter the solenoid valve, and solenoid valves as the opening and closing of water faucets, LCD (liquid crystal display) as output that displays data in the form of letters and numbers; Ethernet shield as output, i.e., connecting a telegram to the internet.

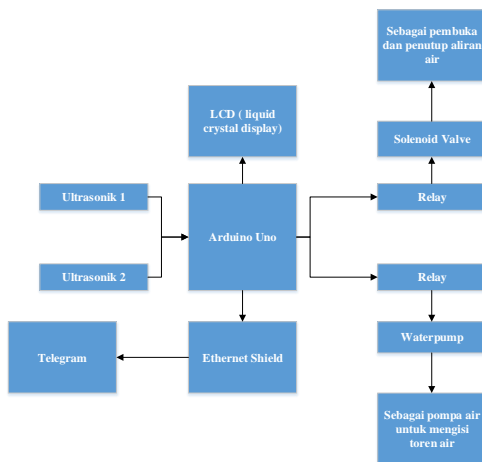


Fig. 3. Block diagram

a) *Series of ultrasonic sensors HC-SR04:* The type of sensor used is an ultrasonic sensor, type HC-SR04. The sensor used to measure the distance of an object is based on the working principle of sound (ultrasonic). The HC-SR04 Ultrasonic Sensor works by shooting sound (ultrasonic) towards objects in front of it. Then the sound is reflected by the object back towards the sensor. The time taken from the sound of the shot until it is received again is used as a parameter to determine the distance between the object and the sensor.

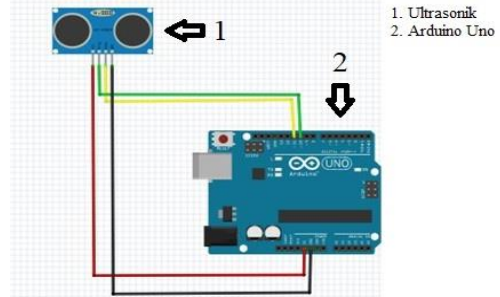


Fig. 4. Ultrasonic sensor circuit

In Fig. 4, this ultrasonic sensor circuit has 4 pins, namely, the GND pin is connected to the GND pin, the VCC pin is connected to the 5V pin, the echo pin is connected to pin 9, and the triangle pin is connected to pin 10.

b) *Relay circuit:* The relay functions as a switch that is operated electrically and carries out the logic functions of the Arduino Uno. The relay is made up of four different components: an electromagnetic coil, an armature, switch contact points (switches), and a spring. The switches located at the contact points themselves are grouped into two types: normally closed (NC) and normally open (NO). What is meant by "normally closed" is the closed position in the initial state when the relay has not been activated. Whereas what is meant by "normally open" is the open position in the initial state when the relay has not been activated.

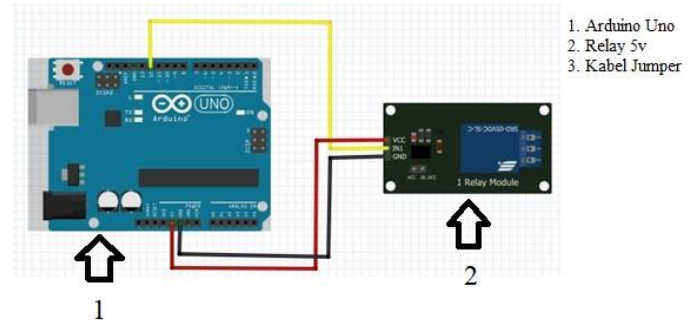


Fig. 5. Series of relay

In Fig. 5, the relay circuit shows that the relay module has 6 pinouts, namely, GND, VCC, IN, NO, NC, and COM. Then the pin is connected to the Arduino, which serves as a connector. Then the GND pin is connected to the GND pin, the VCC pin is connected to the 5 volt pin, and the IN pin is connected to pin 12.

c) *I2C LCD circuit:* The I2C LCD layer is used as an output to display data controlled by Arduino and has a size of 16x2. Because it produces a good display of characters and a lot, LCD is the easiest display medium to observe. On a 16x2 LCD, it can display 32 characters: 16 characters on the top line and 16 characters on the bottom line. Lcd16x2 employs a special driver that allows the 16x2 lcd to be controlled via I2C lines. Through i2c, the LCD can be controlled using only 2 pins, namely SDA and SCL.

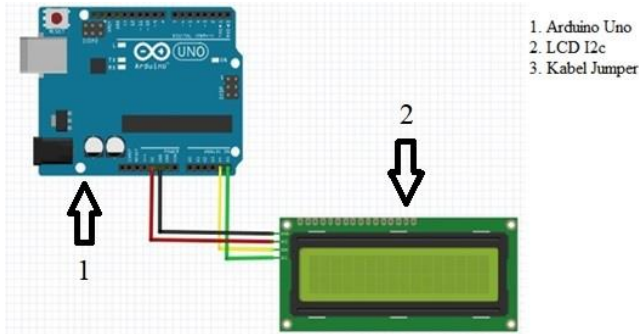


Fig. 6. I2C LCD (liquid crystal display)

Fig. 6 shows that the i2c LCD has 4 pins, namely, the GND pin connected to GND, the VCC pin connected to the 5 volt pin, the SDA pin connected to pin A4, and the SCL pin connected to pin A5.

d) *Relay circuit with solenoid valve and waterpump:* The relay module connects the solenoid valve and water pump to the Arduino. The Arduino's 5 volt pin is connected to the relay module's VCC pin, and the Arduino's GND pin is connected to the relay module's GND pin. Connect the relay module to the positive lead of the 12 volt DC supply and the positive lead of the solenoid valve, and attach the negative lead of the 12 volt DC supply to the negative lead of the solenoid valve. It's the same with the solenoid—how to connect the relay module with the water pump is also the same as with the solenoid. That is the only difference between them and their respective functions.

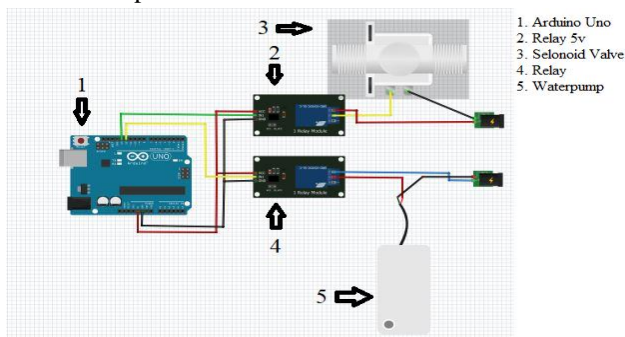


Fig. 7. Relay circuit with solenoid valve and water pump

In Fig. 7, the function of the relay module with the solenoid is to act as a valve control switch on the solenoid. The function of the relay module with the water pump is the on/off switch for the pump controller.

e) *Arduino circuit with Ethernet shield:* In Fig. 8, the last output component is the Ethernet shield. This board is installed directly on the Arduino Uno by stacking it on top. The SPI.h and Ethernet.h libraries manage communication between the Arduino Uno Board and the Ethernet Shield.

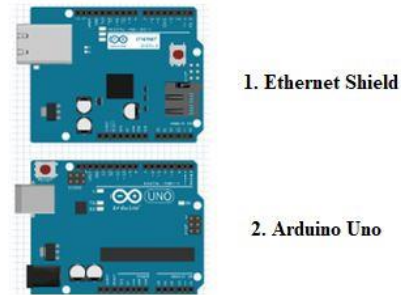


Fig. 8. Arduino uno circuit with Ethernet shield

2) *Network topology design:* The network topology design shows that the Ethernet shield connected to the switch has, of course, been programmed according to the programming instructions that have been made.

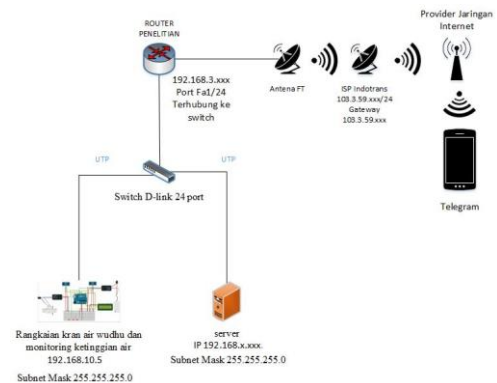


Fig. 9. Network topology design

Fig. 9 is a network design that will be made with a series of tools. To connect the Arduino Uno to the Cisco switch, given the IP address 192.168.x.x./24, use a UTP cable. The CISCO switch is connected to the Mikrotik router with the IP address 192.168.x.xxx/24. Mikrotik router with public IP 103.3.59.xxx connected to ISP Indotrans.

### C. Implementation

In Fig. 10 and 11 at the implementation stage, includes the assembly or installation of all components that are carried out before being implemented in a real system. In addition, the following hardware combinations are implemented in series: HC-SR04 ultrasonic sensor, 5V relay, solenoid valve, water pump, I2C LCD, 12V adapter, Ethernet shield. The following implementation stages will be carried out using a workflow system.

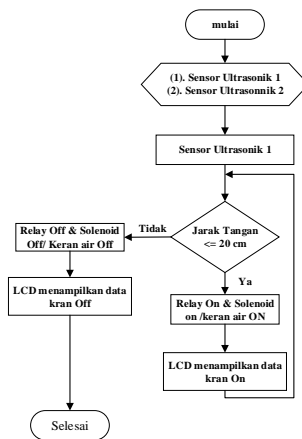


Fig. 10. Water faucet system workflow

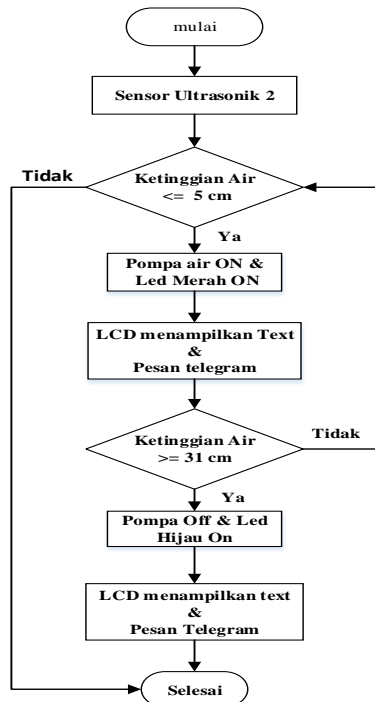


Fig. 11. Water level monitoring system workflow

#### D. Function Testing

At this function-testing stage, namely testing the function of the design tool for the ablation water tap system and monitoring the water tank using Arduino and Telegram, This test is carried out so that the tool is made in accordance with the expected goals.

The following are the stages of testing the function of the design circuit for automatic ablation faucets and water monitoring using Arduino Uno and ultrasonic sensors.

1) *Ultrasonik sensor function testing:* In Fig. 12, at the testing stage of the ultrasonic sensor function, the faucet is a test of the ultrasonic sensor when it detects an object approaching. When the sensor detects an object nearby, the relay will be on, and when the solenoid valve is also on, water will flow.



Fig. 12. Ultrasonik sensor testing

2) *Relay function testing:* In Fig. 13 at this stage, testing the function of the relay is carried out when the relay functions as an on and off switch for the solenoid valve and water pump. When relay one is on, the solenoid valve is on; when relay one is off, the solenoid valve is off; and if relay two is on, the water pump is on; when relay two is off, the water pump is also off.

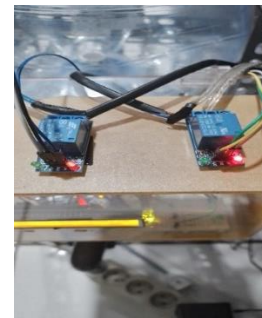


Fig. 13. Relay function testing

3) *Solenoid fungsi solenoid valve:* In Fig. 14, at this stage of testing the function of the solenoid valve, the solenoid valve functions when the relay is on and the solenoid valve is also on. Then, when the relay is off, the solenoid valve is closed. A solenoid valve functions as a substitute for a faucet, whose function is to open and close the flow of water automatically.



Fig. 14. Solenoid valve testing

4) *Waterpump testing:* In Fig. 15, at this stage of testing the function of the water pump, the water pump functions as an increase in water pressure and also as a water filler for the tower. The water pump will turn on if the ultrasonic sensor

detects the water in the tank is low or is about to run out. And the water pump is dead if the water in the tank is full.



Fig. 15. Water pump testing

5) *Ethernet shield testing:* In Fig. 16, this Ethernet shield test, it is done to find out whether the Ethernet shield is connected or not to the internet network so that it can send water level data and control faucets and pumps to see it with the serial monitor on Arduino.

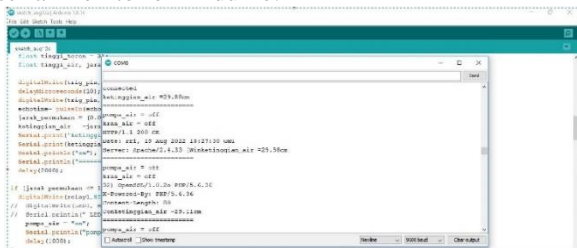


Fig. 16. Ethernet shield function testing

6) *Telegram testing:* In Fig. 17, this stage is carried out to test the function on Telegram. This test is conducted as a notification. This system will send a telegram message in the form of a text on the state of the water tap, water pump, and water level data obtained and will be saved in the database using the abluation water faucet to find out how much content is in the water tank that has been used. The system will send telegram messages in the form of text on the state of the water taps, water pumps, and water level data so that the data obtained will be saved in the database that has been created and can be viewed.



Fig. 17. Telegram testing

7) *Overall tool test results:* The image depicts the state of the abluation water faucet and the monitoring of the water tank with Arduino and Telegram. The ultrasonic sensor that is right at the water faucet functions as a gesture reader. When the hand is under the water faucet with a distance of 20 cm, the relay will be on and the solenoid valve will open the valve so that water comes out automatically when the hand is right under the water faucet. If the distance between the hand and the ultrasonic sensor is greater than or equal to 20 cm, the relay will be off and the solenoid valve will be closed. Furthermore, the ultrasonic sensor located above the water tank functions as a reader of the water level contained in the tank. If the water level in the water tank is less than 5 cm, then the relay is on and the water pump will turn on to fill the water tower. If the water level rises above 30 cm, the water pump will shut down. Then, in the next stage, the system will send a message to the LCD and a telegram in the form of a text of water taps, water pumps, and water level data.



Fig. 18. Overall tool test results

#### IV. CONCLUSION

The conclusions that can be drawn from the results of the tests and discussions that have been carried out regarding the design of the abluation water faucet system and monitoring the water tank using Arduino and telegram. A system has been designed for the application of automatic abluation water taps that can automatically dispense water with the HC-SR04 ultrasonic sensor as a body motion reader and a solenoid valve as a substitute for a faucet that functions as the opening and closing of the water flow. Monitoring the water level in the water tower with the HC-SR04 ultrasonic sensor as a water tap can detect when the contents of the tower are about to run out and alert the level reader in the water tank and a water pump as a water filler can find out the water level in the tank via telegram. After conducting research with the abluation water faucet system and monitoring the water tank using Arduino and telegram, suggestions were made for better development, including 1) In conducting further research, it is expected to be possible to use more than one solenoid valve. 2) It is hoped that further research will monitor via the web and have a database for storing water level data in the tank. 3) For further research, you can expect to use a buzzer when the water in the tank has run out.

REFERENCES

- [1] M. Amin, "Sistem Cerdas Kontrol Kran Air Menggunakan Mikrokontroler Arduino dan Sensor Ultrasonic," *J. Nas. Inform. Dan Teknol. Jar.*, vol. 2, pp. 1–5, 2020.
- [2] H. N. Azhari, "Air Dalam Tafsir Al-Azhar (Kajian Ayat Siklus Air dengan Pendekatan Hidrologi)," Skripsi - Inst. ILMU AL-QUR'AN JAKARTA, 2021, [Online]. Available: [http://27.123.222.2/handle/123456789/1395%0Ahttp://27.123.222.2/bitstream/123456789/1395/3/17210838\\_Publik.pdf](http://27.123.222.2/handle/123456789/1395%0Ahttp://27.123.222.2/bitstream/123456789/1395/3/17210838_Publik.pdf)
- [3] Chrismondari, A. D. Kurniawan, D. Irfan, and Ambiyar, "Dispenser Otomatis Menggunakan Sensor Ultrasonik dan Arduino Uno," *J. Inf. Technol. Comput. Sci.*, vol. 3, no. 2, pp. 227–233, 2020.
- [4] M. Rizal, U. Ungkawa, and M. Premitasari, "Efisiensi Penggunaan Air Pada Sistem Mobil Wudhu," *J. Ilm. Teknol. Infomasi Terap.*, vol. 7, no. 1, pp. 1–8, 2021, doi: 10.33197/jitter.vol7.iss1.2020.444.
- [5] R. Shaputra, "Kran Air Otomatis Pada Tempat Berwudhu Menggunakan Sensor Ultrasonik Berbasis Arduino Uno," *Sigma Tek.*, vol. 2, no. 2, p. 192, 2019, doi: 10.33373/sigma.v2i2.2085.
- [6] D. M. Sani, "KRAN AIR ( Dwikie Mahendra Sani )," 2019.
- [7] J. Cybertech et al., "RANCANG BANGUN ALAT UNTUK MENGUKUR VOLUME TANGKI AIR PADA DEPOT AIR MINUM MENGGUNAKAN TEKNIK SIMPLEX BERBASIS," no. x, 1978.
- [8] Y. A Manurung, "Pembuatan Kran Pencuci Tangan Otomatis dan Hand Dryer Menggunakan Sensor Touch Berbasis Arduino Uno," 2021.
- [9] M. A. Shodiqin and W. D. Kurniawan, "Analisis Sistem Pengendalian dan Pengawasan Level Tangki Air Berbasis Arduino Uno dan Internet of Things," *J. JPTM*, vol. 09, no. 02, pp. 44–53, 2020.
- [10] A. Bintoro and U. Malikussaleh, "SISTEM MONITORING AIR PADA TANGKI BERBASIS INTERNET OF THINGS ( IOT ) BLYNK APP," no. July 2022, 2023.
- [11] D. Nugroho, D. Wahyu Prasetyo, and B. Arifin, "Rancang Bangun Pengendalian Tegangan Menggunakan Kontrol Pid-Arduino Pada Prototipe Mini Plant Mikrohidro," vol. 24, no. 2, pp. 101–115, 2022.
- [12] Wahyu woro lestari, H. Yuana, and S. Kirom, "Rancang Bangun Alat Cuci Tangan Otomatis Berbasis Arduino Untuk Mencegah Penyebaran Covid-19," *JATI (Jurnal Mhs. Tek. Inform.,* vol. 6, no. 1, pp. 38–44, 2021, doi: 10.36040/jati.v6i1.4304.
- [13] A. Rahman Hakim, "Perancangan Dan Implementasi Keran Air Otomatis Dengan Sensor Ultrasonik Berbasis Arduino," *Comasie J. ,* vol. 1, pp. 92–101, 2019.
- [14] U. Y. Syafarida, D. R. Jati, and A. Sulastri, "Analisis Hubungan Konstruksi Sumur Gali dan Sanitasi Lingkungan Terhadap Jumlah Bakteri Coliform Dalam Air Sumur Gali (Studi Kasus: Desa PAL IX, Kecamatan Sungai Kakap)," *J. Ilmu Lingkung.*, vol. 20, no. 3, pp. 437–444, 2022, doi: 10.14710/jil.20.3.437-444.
- [15] faizal Fatturahman and I. Irawan, "Monitoring Filter Pada Tangki Air Menggunakan Sensor Turbidity Berbasis Arduino Mega 2560 Via Sms Gateway," *J. Komputasi*, vol. 7, no. 2, pp. 19–29, 2019, doi: 10.23960/komputasi.v7i2.2422.
- [16] A. Prasetya, "Implementasi Dispenser Pintar Berbasis Internet Of Things Untuk Pemantauan Jumlah Air Minum Guna Menjaga Kesehatan Pada Tubuh Manusia Implementation Of Smart Dispenser Based On Internet Of Things For Monitoring The Amount Of Drinking Water To Maintain He," vol. 8, no. 6, pp. 3026–3035, 2022.
- [17] F.- Puspasari, I.- Fahrurrozi, T. P. Satya, G.- Setyawan, M. R. Al Fauzan, and E. M. D. Admoko, "Sensor Ultrasonik HCSR04 Berbasis Arduino Due Untuk Sistem Monitoring Ketinggian," *J. Fis. dan Apl.*, vol. 15, no. 2, p. 36, 2019, doi: 10.12962/j24604682.v15i2.4393.
- [18] Moch. Bakhrol Ulum, Moch. Lutfi, and Arif Faizin, "OTOMATISASI POMPA AIR MENGGUNAKAN NODEMCU ESP8266 BERBASIS INTERNETOF THINGS (IOT)," *JATI (Jurnal Mhs. Tek. Inform.,* vol. 6, no. 1, pp. 86–93, 2022, doi: 10.36040/jati.v6i1.4583.
- [19] W. Wendanto, D. J. N. Salim, and D. W. T. Putra, "Rancang Bangun Sistem Keamanan Smart Door Lock Menggunakan E-KTP (Elektronik Kartu Tanda Penduduk) Dan Personal Identification Number Berbasis Arduino Mega R3," *Go Infotech J. Ilm. STMIK AUB*, vol. 25, no. 2, p. 133, 2019, doi: 10.36309/goi.v25i2.111.
- [20] F. Bima Prakarsa and Edidas, "Rancang Bangun Alat Sortir Panen Ikan Lele Berbasis Arduino UNO R3," *J. Pendidik. Tambusai*, vol. 6, no. 1, pp. 1202–1218, 2022.
- [21] M. A. Syamsul Arifin, R. Pebriansyah, and B. Santoso, "Prototype Penerapan Internet of Things pada Sistem Informasi Penggunaan Air Rumah Tangga Di BLUD UPT SPAM Kabupaten Musi Rawas," *J. Sustain. J. Has. Penelit. dan Ind. Terap.*, vol. 8, no. 2, pp. 82–90, 2019, doi: 10.31629/sustainable.v8i2.1466.
- [22] D. A. N. Wemos and D. Mini, "PROTOTYPE SISTEM MONITORING DAN PENGENDALI SMART HOME BERBASIS IoT DENGAN MENGGUNAKAN ARDUINO NANO DAN WEMOS D1 MINI TUGAS AKHIR," 2022.
- [23] F. Fitriansyah and Aryadillah, "Penggunaan Telegram Sebagai Media Komunikasi Dalam Pembelajaran Online," *Cakrawala-Jurnal Hum.*, vol. 20, no. 2, pp. 111–117, 2020.
- [24] A. O. Fofid, "Modul Rumah Pintar Berbasis Internet of Thing Dengan Menggunakan Telegram," 2021.

# Risk Analysis of Urban Water Infrastructure Systems in Cauayan City

Rafael J. Padre<sup>1</sup>, Melanie A. Baguio<sup>2</sup>, Edward B. Panganiban<sup>3</sup>, Rudy U. Panganiban<sup>4</sup>, Carluz R. Bautista<sup>5</sup>, Justine Ryan L. Rigates<sup>6</sup>, Allisandra Pauline Mariano<sup>7</sup>  
Isabela State University, Echague, Isabela, Philippines<sup>1, 2, 3, 4</sup>  
Department of Science and Technology, Philippines<sup>5, 6, 7</sup>

**Abstract**—The City of Cauayan Isabela is known as one of the first smart cities and leading agro-industrial centers in the Philippines. Since the center of the economy is in urban areas like Cauayan City, there is a tendency for people and businesses to converge when development and activity take place, with that, a risk analysis was done to analyze hazards for urban water infrastructures in the City of Cauayan. This paper includes an Inventory of the existing urban water infrastructure, with the aid of Geographic Information system Software and gathered data, maps were generated for flood hazards with 5, 25, and 100 yr. return period, liquefaction, ground shaking, and drought of urban water infrastructures. These maps were generated to help the people of Cauayan City, Isabela. The main goal of the paper is to assess the potential prone areas where water infrastructures are located, and monitor areas that are suitable for building such water infrastructures. Problems encountered by the people in utilizing urban water infrastructure can be able to minimize by proper installation of water infrastructures in suitable places which can help the people of the city in water utilization. Since Storm water can cause wide flooding in low elevated areas, to utilize the storm water and to address such problems, an urban water infrastructure with decision support systems intervention can be able to help the city in times of scarcity of water. In addition, the analysis can be used by the local government of the city for proper planning and to project the extent of the hazards.

**Keywords**—Water infrastructure; risk analysis; geographic information systems; decision support systems; storm water

## I. INTRODUCTION

Urban places like Cauayan City in the province of Isabela have a growing population which is a sign that the water demand will increase in future years. Building water infrastructures in places that are prone to hazards like flood, drought, liquefaction, and ground shaking will cause trouble in water utilization which leads to limited sources of water. Thoroughly hazards like flood, drought, liquefaction, and ground shaking will affect the growing economy of the City, and to mitigate the effects of future hazards, a risk analysis can be able to help people to assess the areas where water infrastructures can be installed and for proper planning.

Risk analysis is the process of identifying and assessing potential issues that could negatively impact important business initiatives and operations [1]. This process is utilized in mitigating or reducing certain risks. When performing a risk analysis, adverse events are taken into account, caused by either natural phenomenon, such as severe storms, earthquakes, or floods, or undesirable occurrences brought

about by intentional or unintentional human activities. In addition, the process of a risk analysis helps determine the potential harm from these occurrences, as well as the probability of its occurrences [2].

One of the most frequent types of natural disasters is floods, occurring when an overflow of water submerges land that is usually dry [3]. Floods brought by heavy rainfall can result in a wide range of devastation of critical public health infrastructure, damage of personal property, agricultural sector, and loss of life. From 1998-2017, 2 (two) billion people worldwide were affected by floods [4]. The most vulnerable to floods were the people who live in floodplains or non-resistant buildings, places that are not aware of flooding hazards, or lack warning systems. In this case, a flood risk assessment (FRA) can be done, reviewing the development of documents for its proposal form to consider the possibility of flooding from rivers or groundwater, surface water from sewer sources, estuaries, or even the coast. It must also consider the community and whether a flood risk exists with the development risk to adjacent areas.

Historically, saturated soils have been primarily linked to liquefaction in soils. Unsaturated soils may also be prone to liquefaction in the presence of seismic activity. The consequences of not prioritizing unsaturated soils that are close to saturation as the first rule for liquefaction assessment can be dangerous and disastrous.

Ground shaking is the second main risk for earthquakes due to rapid ground acceleration [5]. There are various levels of ground shaking in one region depending on aspects like topography, type of bedrock, and location and orientation of the fault rupture, all of these have an impact on how seismic waves travel through the ground. Suppose an earthquake is strong enough to cause significant damage to established structures, and sloped terrain may become unstable temporarily or permanently. In a wider extent of earthquakes, districts can be completely destroyed by the effects of ground shaking.

In the natural climate cycle, a drought is a protracted dry period that can happen anywhere. The lack of precipitation makes it a disaster with a slow onset that causes in a shortage of water. Drought can seriously affect agriculture, health, energy, economies, and the environment [6]. Drought affects an estimated 55 million people worldwide every year, and they are the greatest threat to livestock and crops almost everywhere in the world [7]. Due to drought, the livelihood of



individuals is at high risk of disease, death risks are increased and mass migration is fueled. In addition, 40% of the world's population suffers from water scarcity [8], and as a result, the probability of 700 million people being uprooted due to droughts is high by 2030 [7]. On the other hand, regions that are already dry are becoming drier due to rising temperatures brought on by climate change, and wet areas getting wetter. This means that as temperature rises in arid areas, water evaporates more quickly, increasing the possibility of drought or extending the period of drought. Approximately 80-90% of all reported disasters caused by natural calamities over the last ten years have been devastated by floods, drought, tropical cyclones, heat waves, and extreme weather [7].

## II. RELATED WORKS

The identification of locations susceptible to floods and flash floods is an important component of risk management. Floods are natural risk occurrences that vary in severity and cause considerable economic and human losses. They are caused by the interaction of various distinct anthropogenic and natural variables that are particular to a place and have varied impacts on the formation of these events [9]. Around one billion individuals live in flood-prone regions, and floods are regarded as one of the world's most damaging dangers. Under anticipated climate change scenarios, the risks of extreme hydrological events and floods are especially expected to be high and to rise over time [10].

Flash floods are one of the most severe natural disasters, threatening human lives and property in many countries around the world [11] [12]. Floods destroy a large number of people and animals and create catastrophic financial and property damages. They have massive socioeconomic consequences, infrastructural devastation, and environmental disturbance [13]. One of the solutions to solve this is through flood suitability and flood hazard maps that would be useful in assisting local governments, national and international organizations with flood disaster risk reduction and flood shelter design and building [14].

On the other hand, liquefaction is a soil behavior in which strength is reduced and arises due to an increase in pore pressure during earthquake ground shaking on saturated soil [15]. One of the most prevalent seismic consequences that frequently leads to major structure damage during earthquakes is soil liquefaction. Various locations of the world have previously reported liquefaction-induced ground and structure damage in loose, saturated sands and other granular soils [16].

Mapping broad territories for earthquake-induced soil liquefaction danger may appear to be an oxymoron, given that soil liquefaction is a spatially highly limited phenomena in and of itself [17]. In a recent study, they developed combined velocity and fault model that paved the way for further research into seismic segmentation, ground shaking, and rupture modeling [18]. Following the current national earthquake hazard models, the a newly constructed seismogenic source model was established in a paper which includes completely harmonized and cross-border seismogenic sources [19]. In addition, a seismic hazard analysis was also done based from the geologic and geomorphic data [20]. In this study, it includes current and future challenges. Another

study was conducted to develop a region-specific soil behavior type index corrections for evaluating liquefaction hazards [21].

Drought is also considered for assessment in this study. Drought catastrophes endanger agricultural productivity and are projected to worsen as a result of global climate change [22]. Drought analysis was studied that resulted in the identification of key dry periods based on the analyzed drought features, as well as the development of geographic maps of magnitude, length, and intensity for each index for each dry period [23]. Authors have also identified that the standardized precipitation index is used to estimate the drought hazard (SPI) while drought susceptibility is assessed using a variety of indicators, including meteorological conditions, soil characteristics, and irrigation factors [24]. Several models for drought hazards were established like novel hybridized models [25] and MODIS-based Evaporative Stress Index (ESI) and ROC Analysis [26] that offer the spatial resolution required to evaluate regional drought hazard assessment and small-scale agriculture area.

## III. MATERIALS AND METHODS

### A. Flood Hazard

Techniques for assessing the risk of flooding are based on a variety of factors, including meteorological, hydrological, and socioeconomic factors. There are 4 (four) significant phases that are involved in the assessment of flood risk, which include describing the location, estimating the amount of danger, and evaluation of sensitivity and risk as well as intensity. A base map of Cauayan City was obtained from the Local Government Unit of Cauayan, also, secondary data from LiDAR Distribution for Archiving was requested, Using Quantum GIS, this data was processed to determine the extent of flooding. LiDAR flood data includes 5, 25, and 100-yr return periods.

### B. Liquefaction and Ground Shaking Hazard

A base map of Cauayan City was obtained from the Local Government Unit of Cauayan City, also, Secondary data from GEORISK.PH was requested regarding liquefaction and ground shaking, using quantum GIS, this data was processed to determine the extent of liquefaction and ground shaking within the vicinity of Cauayan City.

### C. Drought Hazard

A widely used measure for describing precipitation is the standard precipitation index (SPI) using a variety of timescales for meteorological drought. The SPI is closely related to soil moisture on short time periods, while on longer time scales, it can be related to groundwater and reservoir storage. Regional comparisons of the SPI can be made with climates that differ significantly. It calculates observed precipitation using a consistent scale. Deviation from a chosen probability distribution function represents the raw data on precipitation. Typically, raw precipitation data are fitted to a Pearson type III distribution and then transformed into a normal distribution. SPI values can be interpreted as the number of standard deviations associated with the observed anomaly that deviates from the long-term average. The SPI can be generated using monthly input data for various time periods

ranging from 1 to 36 months.

Rainfall data from PAG-ASA were gathered and served as input to compute the SPI of consecutive months which was analyzed through QGIS. This open-source software was used to analyze and generate a drought hazard map of different existing water infrastructures of Poblacion Cauayan City, Isabela. Existing water infrastructures are: water elevated infrastructures, drainage networks, flood control, and irrigation infrastructures. The principle of the Standardized Precipitation Index was used to analyze and generate maps that include 1, 3, 6, 9 and 12-month SPI through Interpolation. This method requires precipitation data then a calculation of the SPI values out of rainfall data gathered from PAG-ASA was performed to categorize the current level of drought occurring in the City.

#### IV. RESULTS AND DISCUSSION

##### A. Flood Hazard

Flood Hazard data from the LiDAR portal for archiving and distribution was gathered. The local government could use the created map for the Cauayan City government for proper land use planning in flood-prone cones and to identify places at high risk of disaster and manage disaster risk, such as effective and immediate evacuation plans and flooding.

A flood Hazard map with a 5, 25, and 100-year return period was created. Based on the 5-year Flood Hazard Map (see Fig. 1) the Analysis shows that the different Existing water Infrastructures were Classified as Low hazard and Medium Hazard(See Appendix Summary Table for Flood Hazard Map 5-year return period) While on the 25-year Flood Hazard Map (see Fig. 2) the analysis revealed that among the water Infrastructures 9 are classified as medium Hazard, 3 are high risk and the rest are Low Hazard (see Appendix Summary Table for Flood Hazard Map ,25-year return period) and for the 100-year flood hazard Map (see Fig. 3) the analysis shows that among the water Infrastructures, 12 are classified as Medium Hazard, 4 are high risk and the rest of the water infrastructures were low hazard (See Appendix Summary Table for Flood Hazard Map 100-year return period).

##### B. Liquefaction and Ground Shaking Hazard

The assessment was based on the geology and seismic source zone, historical reports of liquefaction, geomorphology, hydrology, and preliminary data from the microtremor survey is used to confirm the type of underlying materials. A semi-detailed map has been developed that can be utilized for land use, emergency response, and mitigation planning but shouldn't be utilized for site-specific evaluation. In addition, no construction is prohibited by liquefaction and ground shaking hazard maps, buildings, and construction in places prone to liquefaction and ground shaking are still possible for as long as the appropriate engineering factors are considered.

Based on the Liquefaction Hazard Map (see Fig. 4) the Analysis shows that the different Existing water Infrastructures were Classified as Low Susceptible, Not Susceptible, and Moderate Susceptible. Among the Water

Infrastructures, 14 are Low Susceptible, 1 – is Moderate Susceptible, and 16 water infrastructures are not Susceptible (see the Appendix Summary Table for Liquefaction Map). In Addition, based on the Ground Shaking Hazard Map (see Fig. 5) based on the revealed analysis it shows that the different Existing water Infrastructures were Classified as Destructive Ground Shaking, labeled as PEIS\* Intensity VII (see Appendix Summary Table for Ground Shaking Hazard Map).

##### C. Drought Hazard

According to the generated 1-month SPI Map (see Fig. 6), among the existing water infrastructures in the different barangays of Poblacion it reflects that it is categorized as near normal which ranges from 0.99 to -0.99 (see Appendix Summary Table of water infrastructures for 1-month SPI Drought Map). The 1-month SPI map depicts a map showing the 30-day period's usual precipitation percentage. However, the generated SPI represents monthly precipitation more accurately because the distribution has been made normal. Based on the generated 3-month SPI map (see Fig. 7) it reflects that all water infrastructures at Poblacion are categorized as near normal which ranges from 0.99 to -0.99. In addition, it appears that some of the barangays were categorized as moderately dry ranging from -1.0 to -1.49 which includes barangay Gappal, Dianao, Manaoag, Linglingay and Buyon (see Appendix Summary Table of water infrastructures for 3-month SPI Drought Map). The 3-month SPI offers a comparison between the precipitation over a certain three-month period and the sum of the 3-month totals of precipitation for each of the years included in the historical records. For the 6-month SPI drought map (see Fig. 8) it reflects that 22 barangays were categorized as moderately dry ranges from -1.0 to -1.49 which includes barangays where water infrastructures located and the rest of barangays in the City were categorized as near normal ranges from 0.99 to -0.99 (see Appendix Summary Table of water infrastructures for 6-month SPI Drought Map). A six-month SPI compares the rainfall for that time frame with the corresponding six-month period over the historical data and can be very effective in showing the precipitation over distinct seasons, While on the generated 9-month SPI map (see Fig. 9), it reflects that existing water infrastructures within the vicinity of Poblacion area of the City were categorized as near normal ranges from 0.99 to -0.99 but based on the map generated (see Fig. 9) there are 9 barangays categorized as moderately dry ranges from -1.0 to -1.49 (See Appendix Summary Table of water infrastructures for 9-month SPI Drought Map). The 9-month SPI shows inter-seasonal precipitation patterns over a medium-term duration, typically, it takes a season or longer for a drought to emerge.

The SPI value below -1.5 for these periods is a good sign that dryness has a major effect on agriculture and might also be having an impact on other sectors. For the 12-month SPI map (see Fig. 10), it also reflects that existing water infrastructures within the Poblacion area were considered as near normal ranges from 0.99 to -0.99 while 17 barangays were categorized as moderately dry ranges from -1.0 to -1.49 (see Appendix Summary Table of water infrastructures for 12-month SPI Drought Map). Long-term precipitation trends are

reflected in the SPI at these timescales. A comparison of the precipitation over 12 consecutive months is referred to as a 12-month SPI which is reported in the same 12 months in a row in every previous year for which data is available. Due to the fact that these timeframes represent the sum of potentially shorter timelines, higher or lower than usual, the longer the SPIs typically converge to zero unless a noticeable dry or wet tendency is present.

TABLE I. APPENDIX SUMMARY TABLE FOR 5-YR FLOOD HAZARD

Appendix Summary Table of Water Infrastructures for 5-yr Flood Hazard Map	
Location	Degree of risk/susceptibility
District 1	Low susceptibility
District 2	Low susceptibility
District 3	Low susceptibility
San Fermin	medium susceptibility
Tagaran	Low susceptibility
Cabaruan	Low susceptibility
Alicaocao	medium susceptibility
Turayong	Low susceptibility
Minante I	Low susceptibility
Minante II	Low susceptibility
Marabulig I	Low susceptibility
Marabulig II	Low susceptibility
Sillawit	Low susceptibility
Alinam	Low susceptibility
Nungnungan I	Low susceptibility
NungnunganII	Low susceptibility
Culalabat	Low susceptibility
Guayabal	Low susceptibility
Baringin norte	Low susceptibility
Buena suerte	Low susceptibility
Rizal	Low susceptibility
Baringin Sur	Low susceptibility
Dabburab	Low susceptibility
San antonio	Low susceptibility
Amobocan	Low susceptibility
San francisco	Low susceptibility
Santa luciana	Low susceptibility
San isidro	Low susceptibility
Naganacan	Low susceptibility
Pinoma	Low susceptibility
Nagrumbuan	Low susceptibility
Labinab	Low susceptibility

Based on the revealed risk analysis with 5-year return period shown in Table I that among thirty-two (32) existing water infrastructures, thirty (30) water infrastructure are low

susceptibility in flood hazard while two (2) water infrastructures labeled as medium susceptibility on flood hazard.

TABLE II. APPENDIX SUMMARY TABLE FOR 25-YR FLOOD HAZARD

Appendix Summary Table of Water Infrastructures for 25-yr Flood Hazard Map	
Location	Degree of risk/susceptibility
District 1	Low susceptibility
District 2	Low susceptibility
District 3	Low susceptibility
San Fermin	medium susceptibility
Tagaran	Low susceptibility
Cabaruan	High susceptibility
Alicaocao	medium susceptibility
Turayong	Low susceptibility
Minante I	Medium susceptibility
Minante II	Low susceptibility
Marabulig I	Medium susceptibility
Marabulig II	Medium susceptibility
Sillawit	Medium susceptibility
Alinam	Medium susceptibility
Nungnungan I	Medium susceptibility
NungnunganII	Medium susceptibility
Culalabat	Low susceptibility
Guayabal	Low susceptibility
Baringin norte	Low susceptibility
Buena suerte	Low susceptibility
Rizal	Low susceptibility
Baringin Sur	High susceptibility
Dabburab	Low susceptibility
San antonio	Low susceptibility
Amobocan	Low susceptibility
San francisco	Low susceptibility
Santa luciana	Low susceptibility
San isidro	Low susceptibility
Naganacan	Low susceptibility
Pinoma	Low susceptibility
Nagrumbuan	Low susceptibility
Labinab	High susceptibility

Based on the revealed risk analysis with 25-year return period shown in Table II that among thirty-two (32) existing water infrastructures, twenty (20) water infrastructure are low

susceptibility in flood hazard and nine (9) medium susceptibility while three (3) water infrastructures labeled as high susceptibility on flood hazard.

TABLE III. APPENDIX SUMMARY TABLE FOR 100-YR FLOOD HAZARD

Appendix Summary Table of Water Infrastructures for 100-yr Flood Hazard Map	
Location	Degree of risk/susceptibility
District 1	Low susceptibility
District 2	Low susceptibility
District 3	Low susceptibility
San Fermin	medium susceptibility
Tagaran	Low susceptibility
Cabaruan	High susceptibility
Alicaocao	High susceptibility
Turayong	High susceptibility
Minante I	Medium susceptibility
Minante II	Medium susceptibility
Marabulig I	Medium susceptibility
Marabulig II	Medium susceptibility
Sillawit	Medium susceptibility
Alinam	Medium susceptibility
Nungnungan I	Medium susceptibility
Nungnungan II	Medium susceptibility
Culalabat	Low susceptibility
Guayabal	Low susceptibility
Baringin norte	Medium susceptibility
Buena suerte	Low susceptibility
Rizal	Low susceptibility
Baringin Sur	High susceptibility
Dabburab	Low susceptibility
San antonio	Medium susceptibility
Amobocan	Low susceptibility
San francisco	Low susceptibility
Santa luciana	Low susceptibility
San isidro	Low susceptibility
Naganacan	Medium susceptibility
Pinoma	Medium susceptibility
Nagrumbuan	Low susceptibility
Labinab	High susceptibility

Based on the revealed risk analysis with 100-year return period shown in Table III that among thirty-two (32) existing water infrastructures, fourteen (14) water infrastructure are

low susceptibility in flood hazard and thirteen (13) medium susceptibility while five (5) water infrastructures labeled as high susceptibility on flood hazard.

TABLE IV. APPENDIX SUMMARY TABLE OF EXISTING WATER INFRASTRUCTURE OF CAUAYAN CITY FOR LIQUEFACTION MAP

Appendix Summary Table of flood control, irrigation and Drainage infrastructures for Liquefaction Map	
Location	Degree of risk/susceptibility
San fermin	Low susceptibility
labinab	Low susceptibility
District I	Low susceptibility
Marabulig I	Low susceptibility
Marabulig II	Low susceptibility
Minante I	Low susceptibility
Minante II	Low susceptibility
Nagrumbuan	Low susceptibility
Pinoma	Low susceptibility
Nungnungan I	Low susceptibility
Nungnungan II	Low susceptibility
Naganacan	Low susceptibility
Alinam	Low susceptibility
Sillawit	Low susceptibility
San Isidro	Low susceptibility & Moderate susceptibility
Santa Lucia	Low susceptibility
San francisco	Low susceptibility & Moderate susceptibility
Amobocan	Low susceptibility & Moderate susceptibility
San Antonio	Low susceptibility & Moderate susceptibility
Dabburab	Moderate susceptibility
Baringin Sur	Moderate susceptibility
Rizal	Moderate susceptibility
Buena Suerte	Low susceptibility & Moderate susceptibility
District III	Not Susceptible
Baringin Norte	Moderate susceptibility
Guayabal	Moderate susceptibility
Culalabat	Moderate susceptibility
Turayong	Not Susceptible
Cabaruan	Not Susceptible
Tagaran	Not Susceptible
Alicaocao	Not Susceptible

Based on the revealed liquefaction risk analysis shown in Table IV that twenty (20) barangays where water infrastructures were installed are classified as low susceptible in liquefaction hazard and six (6) barangays were moderate

susceptibility while five (5) barangays were in the influence of both low and moderate susceptibility on liquefaction hazard.

TABLE V. APPENDIX SUMMARY TABLE OF EXISTING WATER INFRASTRUCTURE OF CAUYAN CITY GROUND SHAKING MAP

Appendix Summary Table of Existing Water Infrastructures for Liquefaction Map	
Location	Degree of risk/susceptibility
San fermin	PEIS* Intensity VII: Destructive Ground Shaking
labinab	PEIS* Intensity VII: Destructive Ground Shaking
District I	PEIS* Intensity VII: Destructive Ground Shaking
Marabulig I	PEIS* Intensity VII: Destructive Ground Shaking
Marabulig II	PEIS* Intensity VII: Destructive Ground Shaking
Minante I	PEIS* Intensity VII: Destructive Ground Shaking
Minante II	PEIS* Intensity VII: Destructive Ground Shaking
Nagrumbuan	PEIS* Intensity VII: Destructive Ground Shaking
Pinoma	PEIS* Intensity VII: Destructive Ground Shaking
Nungnungan I	PEIS* Intensity VII: Destructive Ground Shaking
Nungnungan II	PEIS* Intensity VII: Destructive Ground Shaking
Naganacan	PEIS* Intensity VII: Destructive Ground Shaking
Alinam	PEIS* Intensity VII: Destructive Ground Shaking
Sillawit	PEIS* Intensity VII: Destructive Ground Shaking
San Isidro	PEIS* Intensity VII: Destructive Ground Shaking
Santa Lucia	PEIS* Intensity VII: Destructive Ground Shaking
San francisco	PEIS* Intensity VII: Destructive Ground Shaking
Amobocan	PEIS* Intensity VII: Destructive Ground Shaking
San Antonio	PEIS* Intensity VII: Destructive Ground Shaking
Dabburab	PEIS* Intensity VII: Destructive Ground Shaking
Baringin Sur	PEIS* Intensity VII: Destructive Ground Shaking
Rizal	PEIS* Intensity VII: Destructive Ground Shaking
Buena Suerte	PEIS* Intensity VII: Destructive Ground Shaking
District III	PEIS* Intensity VII: Destructive Ground Shaking
Baringin Norte	PEIS* Intensity VII: Destructive Ground Shaking
Guayabal	PEIS* Intensity VII: Destructive Ground Shaking
Culalabat	PEIS* Intensity VII: Destructive Ground Shaking
Turayong	PEIS* Intensity VII: Destructive Ground Shaking
Cabaruan	PEIS* Intensity VII: Destructive Ground Shaking
Tagaran	PEIS* Intensity VII: Destructive Ground Shaking

Alicaocao	PEIS* Intensity VII: Destructive Ground Shaking
-----------	---

Based on the revealed ground shaking risk analysis shown in Table V, all barangays mentioned were classified as PEIS\* Intensity VII: Destructive Ground Shaking.

TABLE VI. APPENDIX SUMMARY TABLE FOR 1-MONTH SPI MAP

Appendix Summary Table of flood control, irrigation and Drainage infrastructures for 1-month SPI Drought Map	
Location	Description/Value
San fermin	near normal
labinab	near normal
District I	near normal
Marabulig I	near normal
Marabulig II	near normal
Minante I	near normal
Minante II	near normal
Nagrumbuan	near normal
Pinoma	near normal
Nungnungan I	near normal
Nungnungan II	near normal
Naganacan	near normal
Alinam	near normal
Sillawit	near normal
San Isidro	near normal
Santa Lucia	near normal
San francisco	near normal
Amobocan	near normal
San Antonio	near normal
Dabburab	near normal
Baringin Sur	near normal
Rizal	near normal
Buena Suerte	near normal
District III	near normal
Baringin Norte	near normal
Guayabal	near normal
Culalabat	near normal
Turayong	near normal
Cabaruan	near normal
Tagaran	near normal
Alicaocao	near normal

Based on the revealed 1-month SPI drought risk analysis shown in Table VI that all barangays mentioned were classified as near normal for drought hazard.

TABLE VII. APPENDIX SUMMARY TABLE FOR 3-MONTH SPI MAP

Appendix Summary Table of flood control, irrigation and Drainage infrastructures for 3-month SPI Drought Map	
Location	Description/Value
San fermin	near normal
labinab	near normal
District I	near normal
Marabulig I	near normal
Marabulig II	near normal
Minante I	near normal
Minante II	near normal
Nagrumbuan	near normal
Pinoma	near normal
Nungnungan I	near normal
Nungnungan II	near normal
Naganacan	near normal
Alinam	near normal
Sillawit	near normal
San Isidro	near normal
Santa Lucia	near normal
San francisco	near normal
Amobocan	near normal
San Antonio	near normal
Dabburab	near normal
Baringin Sur	near normal
Rizal	near normal
Buena Suerte	near normal
District III	near normal
Baringin Norte	near normal
Guayabal	near normal
Culalabat	near normal
Turayong	near normal
Cabaruan	near normal
Tagaran	near normal
Alicaocao	near normal
Gappal	Moderately Dry
Dianao	Moderately Dry
Manaoag	Moderately Dry
Buyon	Moderately Dry
Linglingay	Moderately Dry

Based on the revealed 3-month SPI drought risk analysis shown in Table VII that thirty-one (31) barangays where water infrastructures were installed are classified as near normal in drought hazard and five (5) barangays were classified as moderately dry on drought hazard.

TABLE VIII. APPENDIX SUMMARY TABLE FOR 6-MONTH SPI MAP

Appendix Summary Table of flood control, irrigation and Drainage infrastructures for 6-month SPI Drought Map	
Location	Description/Value
San fermin	near normal
labinab	near normal
District I	near normal
Marabulig I	near normal
Marabulig II	near normal
Minante I	near normal
Minante II	near normal
Nagrumbuan	near normal
Pinoma	near normal
Nungnungan I	near normal
Nungnungan II	near normal
Naganacan	near normal
Alinam	near normal
Sillawit	near normal
San Isidro	near normal
Santa Lucia	near normal
San francisco	near normal
Amobocan	near normal
San Antonio	near normal
Dabburab	near normal
Baringin Sur	near normal
Rizal	near normal
Buena Suerte	near normal
District III	near normal
Baringin Norte	near normal
Guayabal	near normal
Culalabat	near normal
Turayong	near normal
Cabaruan	near normal
Tagaran	Moderately dry
Alicaocao	near normal
District I	Moderately dry
Mabantad	Moderately dry

Carabatan Chica & Grande	Moderately dry
Gagabutan	Moderately dry
Nagcampegan	Moderately dry
Catalina	Moderately dry
Carabatan bacarena	Moderately dry
Carabatan punta	Moderately dry

Based on the revealed 6-month SPI drought risk analysis shown in Table VIII that thirty (30) barangays where water infrastructures were installed are classified as near normal in drought hazard while nine (9) barangays were classified as moderately dry on drought hazard.

TABLE IX. APPENDIX SUMMARY TABLE FOR 9-MONTH SPI MAP

Appendix Summary Table of flood control, irrigation and Drainage infrastructures for 9-month SPI Drought Map	
Location	Description/Value
San fermin	near normal
labinab	near normal
District I	near normal
Marabulig I	near normal
Marabulig II	near normal
Minante I	near normal
Minante II	near normal
Nagrumbuan	near normal
Pinoma	near normal
Nungnungan I	near normal
Nungnungan II	near normal
Naganacan	near normal
Alinam	near normal
Sillawit	near normal
San Isidro	near normal
Santa Lucia	near normal
San francisco	near normal
Amobocan	near normal
San Antonio	near normal
Dabburab	near normal
Baringin Sur	near normal
Rizal	near normal
Buena Suerte	near normal
District III	near normal
Baringin Norte	near normal
Guayabal	near normal
Culalabat	near normal

Turayong	near normal
Cabaruan	near normal
Parts of Tagaran	Moderately dry
Alicaocao	near normal
Mabantad	Moderately dry
Carabatan chica and grande	Moderately dry
Carabatan punta	Moderately dry
Carabatan bacarena	Moderately dry
Nagcampagan	Moderately dry
Parts of Villa luna	Moderately dry
Parts of Union	Moderately dry
Parts of San luis	Moderately dry

Based on the revealed 9-month SPI drought risk analysis shown in Table IX that thirty (30) barangays where water infrastructures were installed are classified as near normal in drought hazard and the rest of the barangays that are mentioned were classified as moderately dry on drought hazard.

TABLE X. APPENDIX SUMMARY TABLE FOR 12-MONTH SPI MAP

Appendix Summary Table of flood control, irrigation and Drainage infrastructures for 12-month SPI Drought Map	
Location	Description/Value
San fermin	near normal
labinab	near normal
District I	near normal
Marabulig I	near normal
Marabulig II	near normal
Minante I	near normal
Minante II	near normal
Nagrumbuan	near normal
Pinoma	near normal
Nungnungan I	near normal
Nungnungan II	near normal
Naganacan	near normal
Alinam	near normal
Sillawit	near normal
San Isidro	near normal
Santa Lucia	near normal
San francisco	near normal
Amobocan	near normal
San Antonio	near normal
Dabburab	near normal

Baringin Sur	near normal
Rizal	near normal
Buena Suerte	near normal
District III	near normal
Baringin Norte	near normal
Guayabal	near normal
Culalabat	near normal
Turayong	near normal
Cabaruan	near normal
Parts of Tagaran	Moderately dry
Alicaocao	near normal
Mabantad	Moderately dry
Nagcampagan	Moderately dry
Carabatan Chica	Moderately dry
Carabatan Grande	Moderately dry
Parts of Catalina	Moderately dry
Carabatan Bacareno & Punta	Moderately dry
Parts of Villa luna	Moderately dry
Parts of Union	Moderately dry
parts of San luis	Moderately dry
Parts of Gappal	Moderately dry
Manaoag	Moderately dry
Linglingay	Moderately dry
Parts of Buyon	Moderately dry
Parts of Dianao	Moderately dry
Parts of Rogus	Moderately dry

Based on the revealed 12-month SPI drought risk analysis shown in Table X that twenty-nine (29) barangays where water infrastructures were installed are classified as near normal in drought hazard and sixteen (16) barangays that are mentioned were classified as moderately dry on drought hazard.

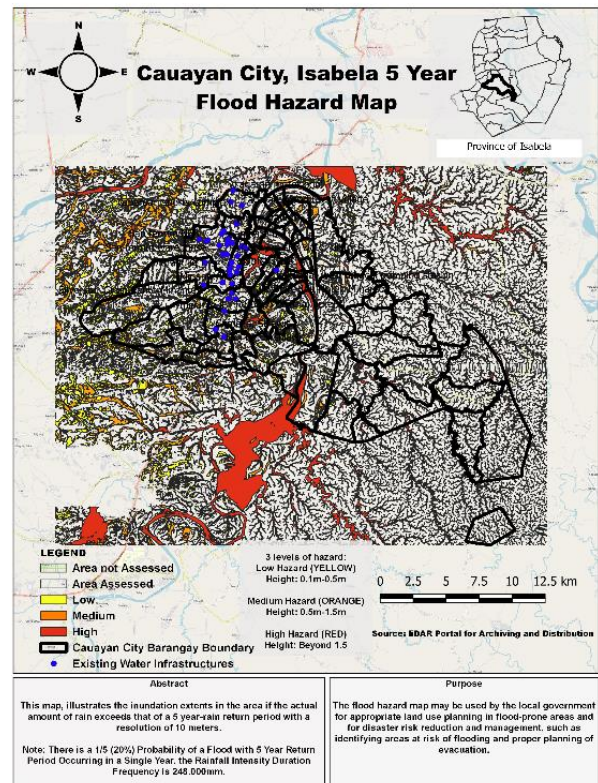


Fig. 1. 5-yr flood hazard map

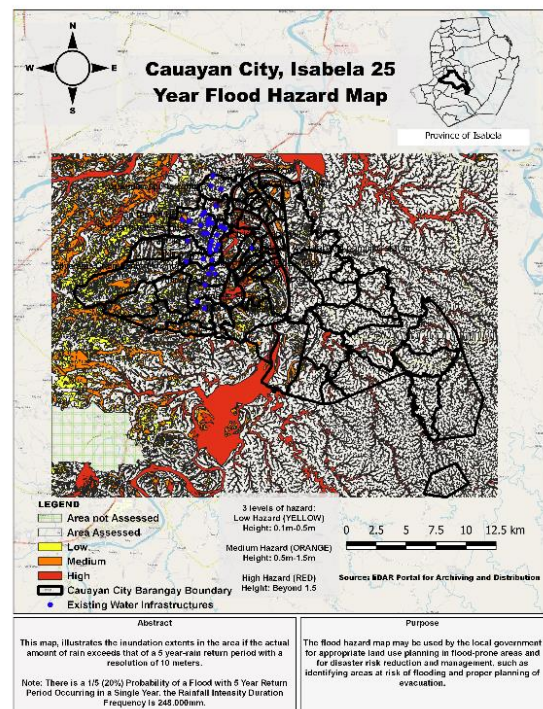


Fig. 2. 25-yr flood hazard map



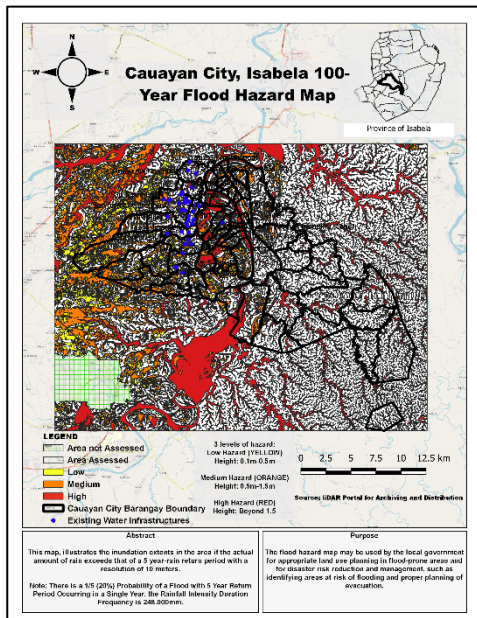


Fig. 3. 100-yr flood hazard map

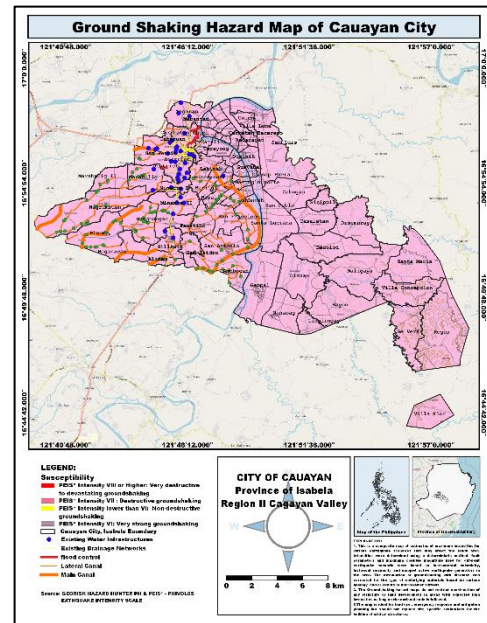


Fig. 5. Ground shaking hazard map

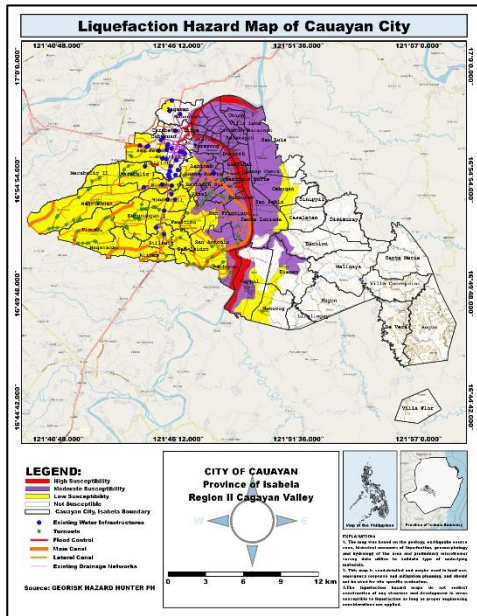


Fig. 4. Liquefaction hazard map

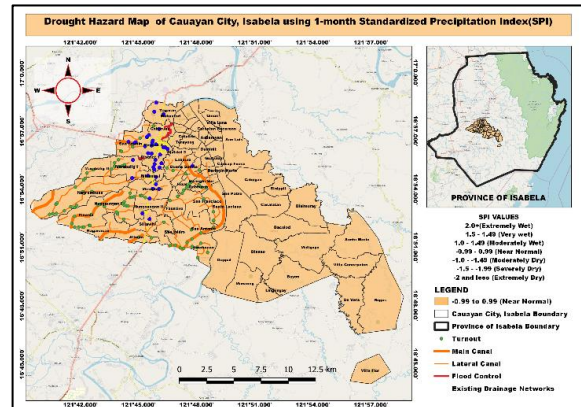


Fig. 6. 1-month SPI map of Cauayan city, Isabela

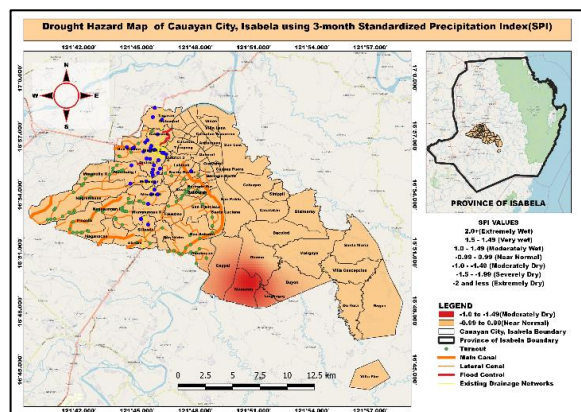


Fig. 7. 3-month SPI map of Cauayan city, Isabela

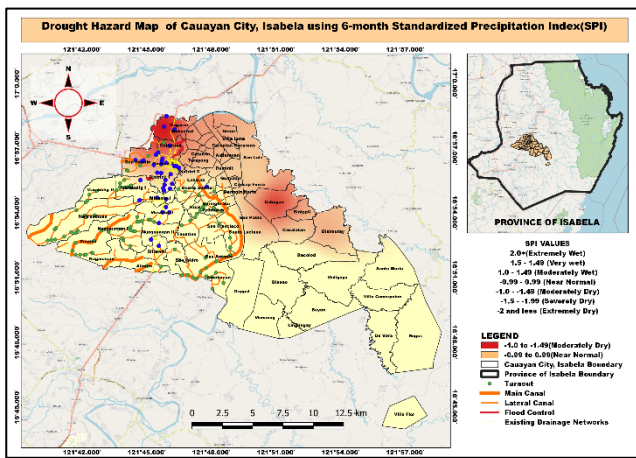


Fig. 8. 6-month SPI map of Cauayan city, Isabela

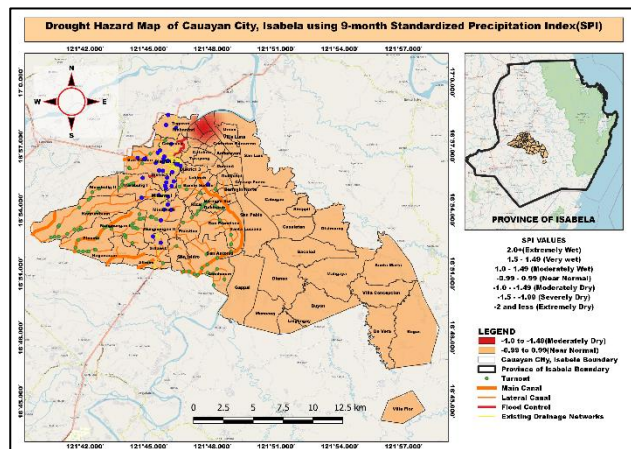


Fig. 9. 9-month SPI map of Cauayan city, Isabela

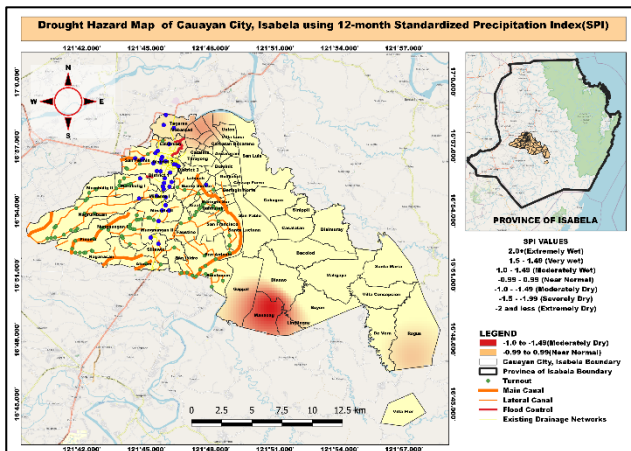


Fig. 10. 12-month SPI map of Cauayan city, Isabela

## V. CONCLUSION AND RECOMMENDATIONS

Risk analysis was done to analyze and identify areas that are prone to different hazards such as flood, Liquefaction, ground shaking, and drought. Based on the flood analysis of existing urban water infrastructures it appears that these infrastructures were at high risk.

The local government would be able to use the generated

hazard maps for identifying flood-prone areas and perform hazard risk reduction and management measures, such as establishing an effective evacuation strategy. In addition, liquefaction, ground shaking, and drought hazard maps also appear the potential areas that are prone to hazards.

## ACKNOWLEDGMENT

The work reported was funded by the Department of Science and Technology (DOST) in collaboration with the Local Government Unit of Cauayan City and Isabela State University (ISU) Echague campus, College of Engineering under Smart Water Infrastructure Management (SWIM) Project 3” GIS-Based Decision Support Tool in Managing Urban Water Infrastructure with Stormwater Intervention for Smart Cities: A Case for Cauayan City.

## REFERENCES

- [1] ACAPS, “Introduction to ACAPS Risk Analysis Methodology,” TAPPI J., vol. 18, no. 5, pp. 1–12, 2019, doi: 10.32964/tj18.5.
- [2] E. B. Panganiban, R. J. Padre, M. A. Baguio, O. B. Francisco, and O. F. Balderama, “An Urban Water Infrastructure Management System Design with Storm Water Intervention for Smart Cities,” Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 4, pp. 306–313, 2022, doi: 10.14569/IJACSA.2022.0130436.
- [3] F. Ibrahim and D. Fritsch, “Flood Spatial Analysis Using LiDAR DEM Study Area ( Azozab locality - Khartoum state - Sudan ),” Int. J. Comput. Sci. Trends Technol., vol. 10, no. 4, pp. 122–133, 2022.
- [4] S. K. Jain, L. Beevers, A. Anandhi, and D. N. Kumar, “Editorial: Flood Management: Multi-Disciplinary Approaches for Data Observation, Analysis, Forecasting, and Management,” Nat. Hazards Earth Syst. Sci., vol. 22, no. 1, pp. 71–83, 2022, doi: 10.5194/nhess-22-71-2022.
- [5] H. Herrmann and H. Bucksch, “Ground Shaking: Ground shaking is the result of rapid ground acceleration,” Dict. Geotech. Eng. Geotech., pp. 639–639, 2014, doi: 10.1007/978-3-642-41714-6\_72471.
- [6] I. R. Orimoloye, “Agricultural Drought and Its Potential Impacts: Enabling Decision-Support for Food Security in Vulnerable Regions,” Front. Sustain. Food Syst., vol. 6, no. February, 2022, doi: 10.3389/fsufs.2022.838824.
- [7] World Health Organization, “Overview Impact WHO response on Drought,” 2022.
- [8] UBS Editorial Team, “Water scarcity impacting 40 % of the world ’ s population,” UBS Glob., pp. 1–5, 2022.
- [9] M. C. Popa, D. Peptenatu, C. C. Draghici, and D. C. Diaconu, “Flood hazard mapping using the flood and Flash-Flood Potential Index in the Buzau River catchment, Romania,” Water (Switzerland), vol. 11, no. 10, 2019, doi: 10.3390/w11102116.
- [10] S. Kittipongvises, A. Phetrak, P. Rattanapun, K. Brundiers, J. L. Buizer, and R. Melnick, “AHP-GIS analysis for flood hazard assessment of the communities nearby the world heritage site on Ayutthaya Island, Thailand,” Int. J. Disaster Risk Reduct., vol. 48, no. November 2019, p. 101612, 2020, doi: 10.1016/j.ijdr.2020.101612.
- [11] A. Alipour, A. Ahmadalipour, and H. Moradkhani, “Assessing flash flood hazard and damages in the southeast United States,” J. Flood Risk Manag., vol. 13, no. 2, pp. 1–17, 2020, doi: 10.1111/jfr3.12605.
- [12] F. S. Hosseini et al., “Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method,” Sci. Total Environ., vol. 711, pp. 1–39, 2020, doi: 10.1016/j.scitotenv.2019.135161.
- [13] S. Janizadeh et al., “Mapping the spatial and temporal variability of flood hazard affected by climate and land-use changes in the future,” J. Environ. Manage., vol. 298, no. August, p. 113551, 2021, doi: 10.1016/j.jenvman.2021.113551.
- [14] K. Uddin and M. A. Matin, “Potential flood hazard zonation and flood shelter suitability mapping for disaster risk mitigation in Bangladesh using geospatial technology,” Prog. Disaster Sci., vol. 11, no. March 2019, p. 100185, 2021, doi: 10.1016/j.pdisas.2021.100185.

- [15] R. Jena, B. Pradhan, M. Almazroui, M. Assiri, and H. J. Park, "Earthquake-induced liquefaction hazard mapping at national-scale in Australia using deep learning techniques," *Geosci. Front.*, vol. 14, no. 1, p. 101460, 2023, doi: 10.1016/j.gsf.2022.101460.
- [16] M. Subedi and I. P. Acharya, "Liquefaction hazard assessment and ground failure probability analysis in the Kathmandu Valley of Nepal," *Geoenvironmental Disasters*, vol. 9, no. 1, pp. 1–17, 2022, doi: 10.1186/s40677-021-00203-0.
- [17] F. Bozzoni, R. Bonì, D. Conca, C. G. Lai, E. Zuccolo, and C. Meisina, "Megazonation of earthquake-induced soil liquefaction hazard in continental Europe," *Bull. Earthq. Eng.*, vol. 19, no. 10, pp. 4059–4082, 2021, doi: 10.1007/s10518-020-01008-6.
- [18] P. Bürgi, J. Hubbard, S. H. Akhter, and D. E. Peterson, "Geometry of the Décollement Below Eastern Bangladesh and Implications for Seismic Hazard," *J. Geophys. Res. Solid Earth*, vol. 126, no. 8, 2021, doi: 10.1029/2020JB021519.
- [19] L. Danciu et al., "The 2020 update of the European Seismic Hazard Model: Model Overview," no. EFEHR Technical Report 001 v1.0.0, pp. 1–121, 2021.
- [20] K. D. Morell, R. Styron, M. Stirling, J. Griffin, R. Archuleta, and T. Onur, "Seismic Hazard Analyses From Geologic and Geomorphic Data: Current and Future Challenges," *Tectonics*, vol. 39, no. 10, pp. 1–47, 2020, doi: 10.1029/2018TC005365.
- [21] B. W. Maurer, R. A. Green, S. van Ballegooy, and L. Wotherspoon, "Development of region-specific soil behavior type index correlations for evaluating liquefaction hazard in Christchurch, New Zealand," *Soil Dyn. Earthq. Eng.*, vol. 117, no. April 2018, pp. 96–105, 2019, doi: 10.1016/j.soildyn.2018.04.059.
- [22] Z. Zeng, W. Wu, Z. Li, Y. Zhou, Y. Guo, and H. Huang, "Agricultural drought risk assessment in Southwest China," *Water (Switzerland)*, vol. 11, no. 5, pp. 1–20, 2019, doi: 10.3390/w11051064.
- [23] G. R. Singh, M. K. Jain, and V. Gupta, "Spatiotemporal assessment of drought hazard, vulnerability and risk in the Krishna River basin, India," *Nat. Hazards*, vol. 99, no. 2, pp. 611–635, 2019, doi: 10.1007/s11069-019-03762-6.
- [24] L. Zhang, W. Song, and W. Song, "Assessment of agricultural drought risk in the Lancang-Mekong Region, South East Asia," *Int. J. Environ. Res. Public Health*, vol. 17, no. 17, pp. 1–24, 2020, doi: 10.3390/ijerph17176153.
- [25] O. Rahmati et al., "Capability and robustness of novel hybridized models used for drought hazard modeling in southeast Queensland, Australia," *Sci. Total Environ.*, vol. 718, p. 134656, 2020, doi: 10.1016/j.scitotenv.2019.134656.
- [26] D. Hazard, M. E. Stress, I. Esi, and R. O. C. Analysis, "Drought Hazard Assessment using MODIS-based Evaporative Stress Index (ESI) and ROC Analysis," vol. 62, no. 3, pp. 51–61, 2020.

# Mitigate Volumetric DDoS Attack using Machine Learning Algorithm in SDN based IoT Network Environment

Kumar J<sup>1</sup>, Dr Arul Leena Rose P J<sup>2\*</sup>

Research Scholar<sup>1</sup>, Associate Professor<sup>2</sup>

Department of Computer Science-College of Science and Humanities,  
SRM Institute of Science and Technology, Chengalpattu, India<sup>1,2</sup>

**Abstract**—Software-Defined Networking (SDN) is a recent trend that is combined with Internet of Things (IoT) in wireless network applications. SDN focus entirely on the upper-level network management and IoT enables monitoring the physical activity of the real-time environment via internet network connectivity. The IoT clusters with SDN often undergoes challenges like network security concerns like getting attacked by a Distributed Denial of Service (DDoS). The mitigation of network management issues is carried out by the frequent software update of SDN. On other hand, the security enhancement is needed to all alleviate the mitigation of security attacks in the network. With such motivation, the research uses machine learning based intrusion detection system to mitigate the DDoS attack in SDN-IoT network. The control layer in the SDN is responsible for the prevention of attacks in IoT network using a strong Intrusion Detection System (IDS) framework. The IDS enables a higher-level attack resistance to the DDoS attack as the framework involves feature selection-based classification model. The simulation is conducted to test the efficacy of the model against various levels of DDoS attacks. The results of simulation show that the proposed method achieves better classification of attacks in the network than other methods.

**Keywords**—DDoS; SDN; IoT; machine learning

## I. INTRODUCTION

As the frequency of cyberattacks on governments and corporations across the world rises, academics like have been working feverishly to develop effective network intrusion detection systems (NIDS) [1].

Web-based platform attacks, DoS attacks, and malicious insiders are among the most severe forms of cybercrime [2]. Businesses risk having their intellectual property stolen, and governments risk having interruptions to their key national infrastructure if malicious software is permitted to infiltrate the system. Companies use security measures like firewalls, antivirus programs, and NIDS to keep hackers out of their networks [3].

A new architecture, the software-defined network (SDN) [4] divides the network control functions from its forwarding functions. By physically separating the control plane from the data plane, it is feasible to facilitate uncomplicated management of the network [5]. This component of SDN makes it easier to create new types of applications, which in turn promotes the need for a new sort of networking paradigm

that can handle NIDS. Using SDN, developers may quickly and easily design innovative software. SDN controllers have the capacity to incorporate machine learning and deep learning (ML/DL) methodologies [6]. This facilitates better network visibility and security when IoT is interfaced with SDN.

Earlier attempts to deploy NIDS utilizing SDN controllers and deep learning techniques have met with mixed success. In [6], the authors implement a controller based on an anomalous algorithm. They constructed a deep neural network in order to cut down on the number of distinguishing criteria that may be used to discern regular traffic from abnormal traffic. In addition to this, they utilized deep learning strategies in order to evaluate their model [7].

From the problems stated above, it is found that there exists a gap of concurrent processing that effectively retain the level of accuracy while mitigating the attacks in the network.

The selection of a lightweight attack mitigation mechanism using a machine learning (ML) can pose a lighter load to the network and may not affect the network in terms of its computational burden.

In this paper, the research uses machine learning based intrusion detection system to mitigate the Distributed Denial of Service (DDoS) attack in SDN-IoT network. The Intrusion Detection System (IDS) enables a higher-level attack resistance to the DDoS attack as the framework involves feature selection-based classification model.

## II. RELATED WORKS

In order to defend the control and data planes from DDoS attacks, Shoeb and Chithralekha [8] established a controller process priorities are set according to the node trust level, which is configured based on the node behaviour during regular business hours. The node worth is calculated based on its activity. In high-demand situations, the controller is set up to ignore requests from some nodes if the sum of their requests has already reached a predetermined maximum. The controller makes a rule change to one with a shorter timeout, prompting a response from the standard nodes as well.

Support vector machine (SVM) classifiers are proposed by Kokila et al. [9] for detecting DDoS attacks. The SVM must be trained with historical data before it can reliably predict the behaviour of unobserved traffic samples. When compared to

other simulated methods, the SVM has superior accuracy and fewer false positives. However, SVM is very dependent on the accuracy of the data used to train the model.

Xiao et al. [10] suggest a concept that employs a bloom filter in the SDN to detect link flooding attack. When the collector detects a deviation in link use, it checks the switch flow table to determine whether any aberrant flows can be deduced from the data. The detector can perform packet sniffing thanks to a controller that watches the network in real time. Seeing as the Bloom filter retains crucial IP features, it can be used to determine if the packet current classification is anomalous. However, neither a definition of anomalous link consumption nor the controller detection mechanisms are provided.

Lim et al. [11] suggest that changing one IP address is one way to lessen the impact of distributed denial of service attacks. If a host changes its IP address but then continues to send more than some threshold number of packets to the old address, the host will be marked as a bot and banned. It appears from the simulation results that bot-driven DDoS attacks can be countered. The question of what metric or threshold should be used to initiate defensive measures and withdrawal from a conflict remains unclear.

Similar to SVM and Self Organising Machine (SOM), the method for categorizing DDoS attacks described by Phan et al. [12] combines the two models. Both SVM and SOM models can be trained and tested using pre-existing data sets. Each protocol employs its own distinct set of SVMs to filter control-plane communications. If the SVM determines that data from a certain flow could be associated with the attack region, it will forward that data to the classifier. The SOM must decide if the current is appropriate for the current style era. Simulations show that when SVM and SOM are used together, better results are achieved than when using either approach alone.

An approach to detecting distributed denial of service (DDoS) attacks is presented by Chin et al. [13], which makes use of the interplay between a monitor, a correlater, and a controller. The Monitor component alerts the IDS whenever it finds something out of the ordinary on the network. Once the IDS has confirmed the existence of an attack, it will send the relevant data to the correlater.

Hameed and Khan [14] developed a secure protocol using a cooperative DDoS mitigation technique. The exchange of messages, certificates, and signatures are the fundamental

building blocks of the C-to-C protocol. While the signature component is responsible for ensuring the accuracy and integrity of the data, the certificate component is responsible for establishing a trust connection between authenticating controllers. When a controller detects a DDoS attack, it warns its surrounding controllers as soon as possible by transmitting a list of malicious IP addresses and by modifying the policy on the data plane. As a result, access to these packets is being prohibited across the network in a number of different locations. The results of the simulation illustrate how rapidly this method may warn neighbouring controllers and thwart attacks.

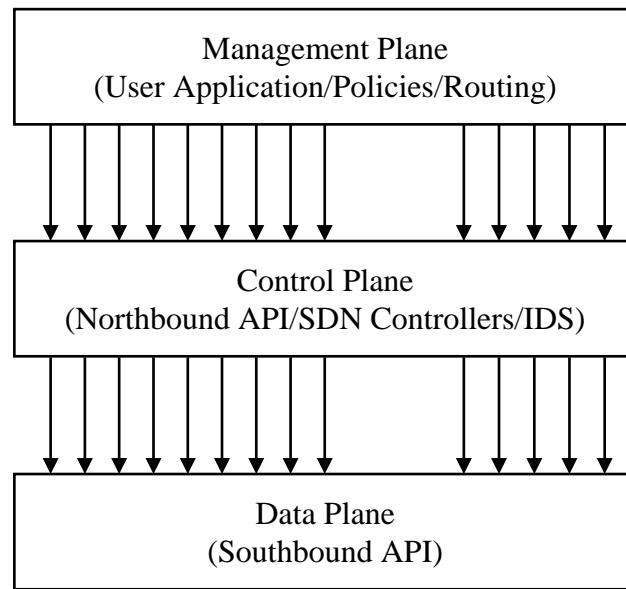
Macedo et al. [15] came up with the idea of a multi-controller cluster when they were developing The DDoS Attack Mitigation in SDN Networks is comprised of these three stages: (i) identifying the overloaded controller by using control message latency or stability; (ii) selecting the controller with the highest performance for coordinated mitigation; and (iii) reducing the impact of attacks. The model identifies the overloaded controller by using control message latency or stability.

A DDoS security architecture was introduced by Sahay et al. [16]. Its primary goal is to reduce the amount of destructive Internet traffic. In order to identify network flows, and the customer end detection engine is the one that determines whether or not the traffic flow is malicious. The status of the connection is communicated from the controller belonging to the customer to the controller belonging to the service provider. A determination made by the ISP controller directs that the harmful flow be sent to the filter so that it can be examined in further detail. Nevertheless, the communication between controllers also needs to be secure.

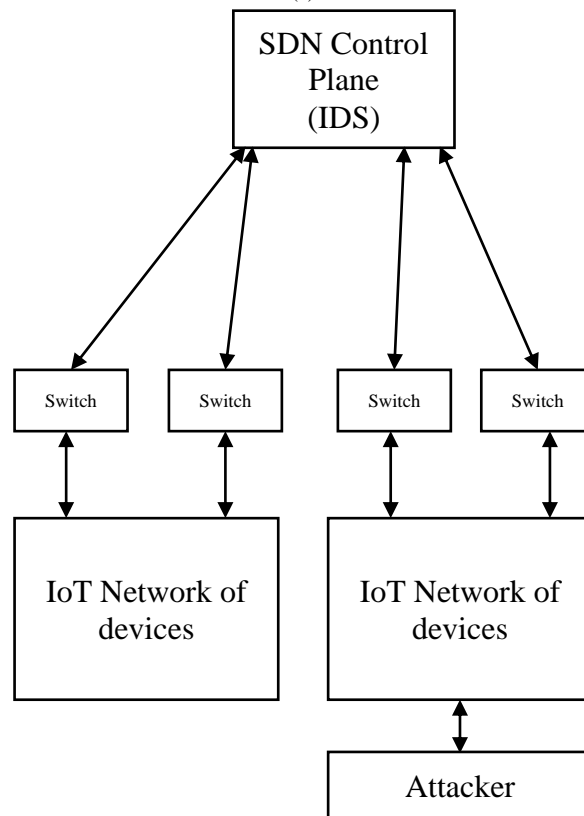
### III. PROPOSED METHOD

The research makes use of an intrusion detection system that is founded on machine learning in order to mitigate the impact that DDoS attacks have on SDN-IoT networks. The layers of SDN are provided in Fig. 1(a) and when the IoT network connected with SDN layers under attacks is illustrated in Fig. 1(b).

In Fig. 1(a) and 1(b), the control plane in SDN is responsible for mitigating the attacks in IoT network as it is embedded with a strong IDS framework that reads the network logs and classifies the traffic, and mitigates the attacks based on the anomalies present.



(a)



(b)

Fig. 1. SDN layered architecture b) IoT network with SDN architecture

It employs a feature selection-based classification methodology, the IDS is able to withstand more complex varieties of DDoS attack. A simulation is carried out in order to check that the model is capable of withstanding DDoS attacks of varying degrees of severity. The findings of the simulation reveal that the suggested strategy works better than competing alternatives when it comes to classifying network intrusions.

In Fig. 2, a variety of strategies and processes are illustrated for the purpose of putting an intrusion detection system into action. A number of methods have been created, and they can be roughly categorized as statistical approaches, data mining techniques, and machine learning-based methods. These methods were designed to identify anomalies in the data.

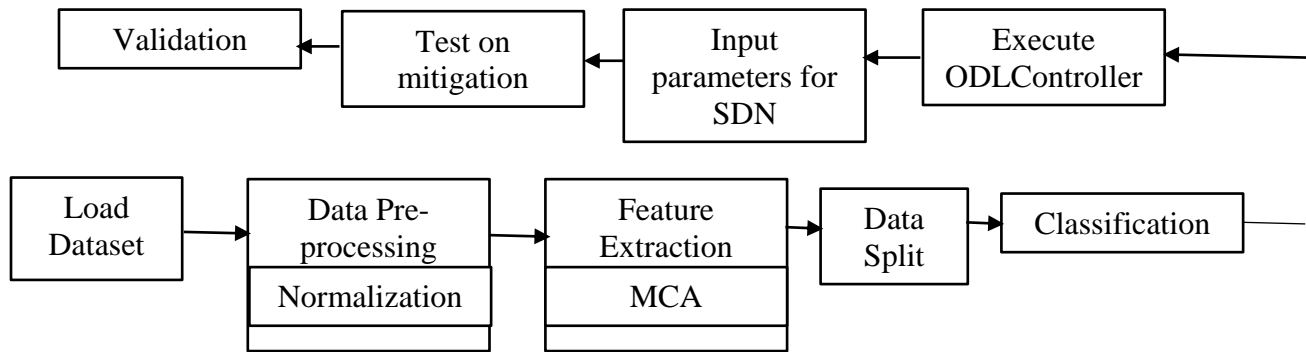


Fig. 2. Proposed IDS framework in SDN controller

A NIDS that relies on signatures can only detect previously discovered strains of malicious software. The detection system makes use of a set of rules that are derived from a combination of packet header and packet content inspection utilizing a present signature in order to identify potentially malicious network activity.

An ML/DL-based SDN-based intrusion detection system excels in a number of different areas, including security enforcement, virtual management, and QoS. SDN provides us with the opportunity to enhance the security of our networks by doing away with the requirement for specialist hardware, providing us with greater creative leeway in the manner in which we build our networks, and making it simpler for us to put into action new security measures.

A SDN can be constructed with fully adjustable features and software implementations of switches by making use of platforms that simulate and emulate real-world conditions. When it comes to putting the ideas of SDN into practice, Open Flow is one of the protocol standards that have been embraced by the largest number of organizations.

The SDN controller, also referred to as a network operating system, is an essential component of SDN networks. The SDN controller is responsible for presenting a consolidated image of the network as well as coordinating interactions with all programmable network components. In addition, there are already a variety of alternative SDN controllers that can be utilized. Fig. 4 is a representation of a network intrusion detection system that is based on SDN.

When compared to NIDS, machine learning makes it abundantly evident that researchers have started applying deep learning techniques. Deep learning is a potential strategy for the next generation of intrusion detection methods, as it can automatically uncover a correlation in the data. When applied to a wide variety of categorization difficulties in SDN networks, methodologies based on ML performed significantly better than state-of-art methods.

Classification problems appear to be where supervised machine learning systems shine the brightest. ML-based approaches performed far better than classic machine learning methods did. Due to the fact that the characteristics of attacks are unknown, unsupervised learning algorithms such as stacked autoencoders, RNNs, and hybrid-based algorithms will prove to be the most successful algorithms for implementing NIDS on an SDN platform.

### A. Pre-Processing

The technique begins with the normalization of the dataset. The raw data is processed so it is standardized for the subsequent processing steps. This makes the entire thing simpler to design and more effective to put into practice. The first normalization of both the training/testing data set happens during the pre-processing step of the procedure.

The degree to which the underlying data are standardized determines a significant portion of the success of the weight coefficients selection approach. The standardization strategy does not incorporate any approaches that can be used to manipulate the detection rate in any way. Normalization of the measured values on a separate scale is conceptually shared scale before averaging the results of the measurements. There are many different kinds of normalization, and some of them require a rescaling technique in order to obtain values that are associated with an entirely new variable. The equation for the normalization of the mean and standard deviation is given as below:

$$\sigma_i^2 = -\frac{1}{(1-m-n)} \sum_{j=1}^n \varepsilon_j^2 \quad (1)$$

where

$\sigma$  - standard deviation

$m, n$  - parameter.

When the errors can be distinguished from one another, they can be formulated as follows:

$$g_i \sim \frac{T\sqrt{O}}{\sqrt{O+T^2-1}} \quad (2)$$

where

$T, O$  – initialized datasets and

$g$  - random variable.

The standard deviation is employed to analyze the change in the variable.

$$K = \frac{\mu^k}{O^k} \quad (3)$$

where

k - moment scale,

$\mu$  - normally ordered distribution

$$\mu^k = S(X - \mu)^k \quad (4)$$

where

X - random variable and

s - expected value

$$O^k = \left( \sqrt{s(X - \mu)^k} \right)^a \quad (5)$$

The scaling is used to normalize the normalizing the variable's distribution,

$$s_v = \frac{S}{X} \quad (6)$$

where

$s_v$  - coefficient of variance.

The study determines the adjusted new normalized value. When applied in mapping fashion, the resulting value can take on values between 0 and 1. Standardization helps give a better training instance since it ensures that all of the training data shares the same field, which could range between 0 and 1. The normalizing formula is represented as below:

$$Ss = (S - S_{min}) / (S_{max} - S_{min}) \quad (7)$$

where

$S_s$  - output of the normalization procedure and S - initial value.

$S_{max}$  and  $S_{min}$  - maximum and minimum attribute values, respectively.

The main aim of the pre-processing involves the data standardization to eliminate the restrict the duplicate information and missing statements.

During this stage of pre-processing, both the input IDS datasets and the dataset itself are standardized, and the dataset is also normalized using the missing data.

1) *Feature extraction:* Multilinear Component Analysis (MCA) is widely used for the extraction of most relevant feature. This is because MCA is able to extract the most nuanced features through the incorporation of these search methods. The MCA places a high value on both the redundancy that exists between features and the unique extrapolative potential that each one possesses. There are many different ways to obtain attribute information.

$$Y = V\Sigma U^T \quad (8)$$

where

$V \in R^{S \times S}$  - column orthogonal matrices of Y,

$U \in R^{M \times M}$  - row orthogonal matrices of Y.

$\Sigma$  - diagonal matrix

Thus an attribute obtained using the attribute function (Y) is the combination of orthogonal matrix of both rows and columns and a diagonal matrix.

The capacity of a component to supply information for the data it is believed that the image is mirrored in the variation of the projection. The following equation can be used to evaluate the performance of feature extraction:

$$C_{xc} = \frac{tc_x^j}{\sqrt{t + t(t-1)c_j^j}} \quad (9)$$

where

$C_{xc}$  - correlation between the variables and subsets,

n - attributes.

$c_x^j$  - correlation between the variable and attributes.

$c_j^j$  - average inter-correlation between the attributes.

The ratio of principal components is defined as below:

$$C = \frac{\sum_{n=1}^P \eta_n^2}{\sum_{n=1}^N \eta_n^2} \quad (10)$$

A few principal components can preserve more than 90% of the overall variance of the Y data. This is the case even though the CCV ratio is calculated using only the variables.

### B. Attack Detection

A model distinguishes between a huge volume of normal traffic in order to protect itself from a decentralized attack. The specified input is utilized by the attack in order to perform an estimation of the discrete scalable memory-based attack vector probability approach. Estimating the value of a random variable requires taking its distribution into account as the starting point of the process. The primary goals are to discover context and keep a close eye on relevant data as it emerges by chance. Only by persistent, day-to-day effort will it be possible to realize the goal. It is possible to have a look at the user tendencies as well as the value that is at risk. Using SVM, one may determine the likelihood of a vector value being one of several possible values. It is found by multiplying the value of the standard deviation by the constant that is used in the calculation.

$$P\left(\frac{\emptyset}{x}\right) = \frac{\left(\frac{x}{\emptyset}\right)P(\emptyset)}{P(x)} \quad (11)$$

where



$P(\emptyset)$  - probability distribution function,

$P(x|\emptyset)$  - likelihood function.

$P(\emptyset/x)$  - evidence function.

The probability distribution provides the value of both likelihood and evidence function to find the rate at which the vector value changes.

A direct connection may be made between the likelihood and the posterior probability. The likelihood of the probability is defined as below:

$$F(x) = \frac{f(x)L\left(\frac{x}{y} = y(x)\right)}{\int_{-\infty}^{\infty} f(x)(u)L\left(\frac{x}{y} = y(x)\right)(u)+u} \quad (12)$$

where

$F(x)$  - prior density function.

$f(x)L\left(\frac{x}{y} = y(x)\right)$  - likelihood function.

$f(x)(u)L\left(\frac{x}{y} = y(x)\right)$  - normalizing constant.

Using the probability equation, we were able to essentially create a map of the function irregularity. Following the computation of the attack baseline probability, the results were presented. After that one is able to evaluate the significance of the attack path.

$$A = \min_{i=1:M} OS_i^* \quad (13)$$

After determining the extent of the damage caused by an attack, it is possible to classify the attack. The vector technique might make use of convolutional layers as a crucial layer in order to acquire knowledge more effectively from the input data. It does what its job title implies and reduces low-level features (kernels). However, the convolutional operator lacks rotation invariance as a property of its own. In addition, additional layers in the stack are pooling layers, which result in a reduction in the amount of data. In fact, neural networks and other fully connected structures form the core of the suggested classifier architecture. The first stage of training a classifier is called the feed phase, and the second stage is called the reverse propagation phase.

The network error is what is utilized to generate the parameter gradient, and with that, the weight matrices are updated as part of the process of context propagation. This is

all done by utilizing the network error. Classifier-trained systems must be governed by large amounts of data in order to successfully accomplish classification jobs. Because the classification error is reduced in proportion to the depth of the classifier, which it is carried out.

The target is given a score based on the probabilities that are calculated for it. It is possible to calculate the difference between a single variable and a number of other variables by employing a technique known as the SVM. The recommended method can initially read and reorganize the data and then evaluate the identified technique based on the likelihood that it belongs to each class.

$$F = \Delta M - k(A(M))^2 \quad (14)$$

where

$F$  - feature,

$M$  - pointed feature,

$\lambda_1\lambda_2$  - classified features.

$$\Delta M = \lambda_1\lambda_2$$

$$A(M) = \lambda_1 + \lambda_2$$

The classification based on the scalable memory is defined as below:

$$F = \lambda_1\lambda_2 - K(\lambda_1 + \lambda_2)^2 \quad (15)$$

where

$K$  - empirical constant.

### C. Attack Mitigation

The SDN controllers are able to instruct switches regarding the destination to which packets should be sent when they make use of the Open Flow Protocol Specification. Researchers are able to conduct experiments on real-world networks thanks to a protocol standard known as Open Flow. This protocol standard details the message formats that are utilized by each controller.

This technique needs the establishment of a robust network in addition to the expensive acquisition of various assets so that it can withstand an attack. In the event of a serious DDoS, the upstream network resources need to be severed in order to preclude any reliable local response. Before subscribing to a DDoS mitigation service, there are a number of elements that need to be taken into consideration. Some of these factors include scalability, endurance, stability, and network traffic. This is only a passing phase that will resolve itself when some time has passed. The modified attack device packets will be transmitted to an analysis service at some point in time.

#### IV. RESULTS AND DISCUSSIONS

In this section, setting up the cloud with the necessary conditions exist is part of the system model configuration. Using the deep learning model, we report the results obtained from the CTU-13 botnet and the ISCX 2012 IDS datasets. Ninety per cent of the total sample size is split evenly between the training and testing stages. To clarify, this means that 90% of the sample is used for training, while the remaining 10% is split evenly between validation and test data. Our group used a 10-cross validation technique to ensure the accuracy of the results. The current plan is to use 9 partitions from the total data samples as training samples and 1 partition as a testing sample. Each of them will be selected at random. The procedure is repeated ten times, and the final result is the mean of these ten separate estimates.

The proposed method is compared with existing machine learning methods like ANN, SVM and SOM.

TABLE I. SIMULATION PARAMETERS

Parameters	Values
Packet Size	1442byte
Bandwidth	10 Mbps
Monitoring Time	20 s
Forwarding Method	Fitness route

The 99.12% accuracy achieved on the training data sample is matched by the 98.88% accuracy achieved on the test sample using Table I.

Fig. 3 shows the results of accuracy in detecting the attacks with the conventional NIDS models. The results show that the proposed NIDS has a higher rate of accuracy than the other methods. The range of accuracy using the proposed SVM classifier along with the feature extraction and pre-processing information is between the range of 95-96% and this is higher than the existing methods.

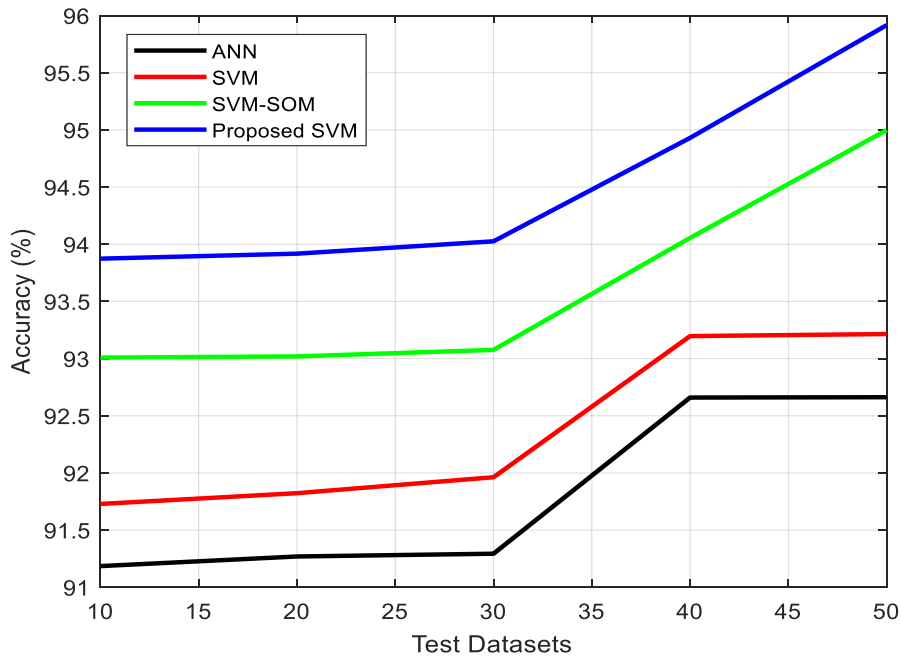


Fig. 3. Accuracy

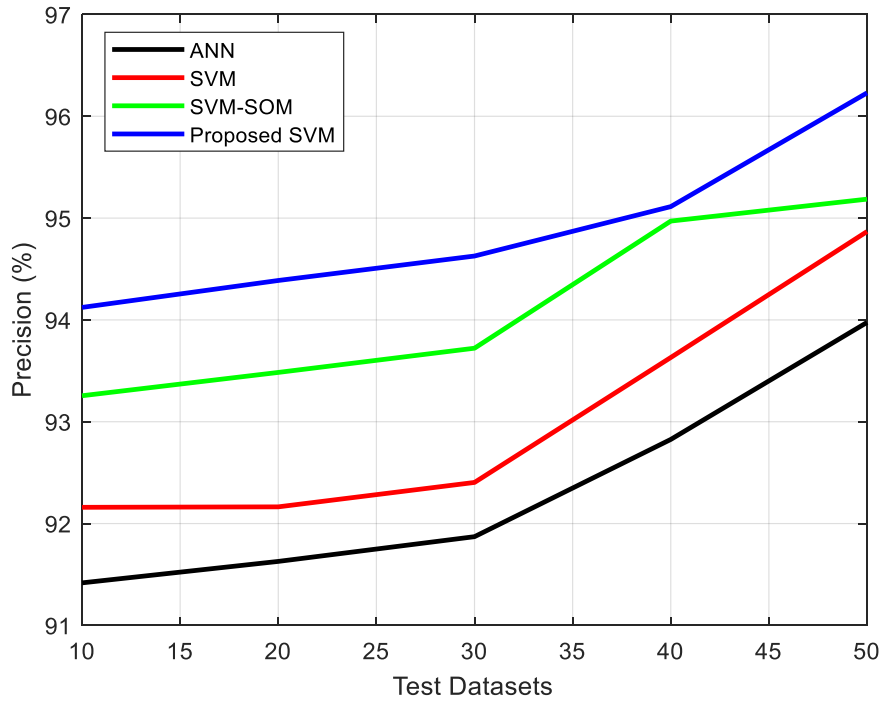


Fig. 4. Precision

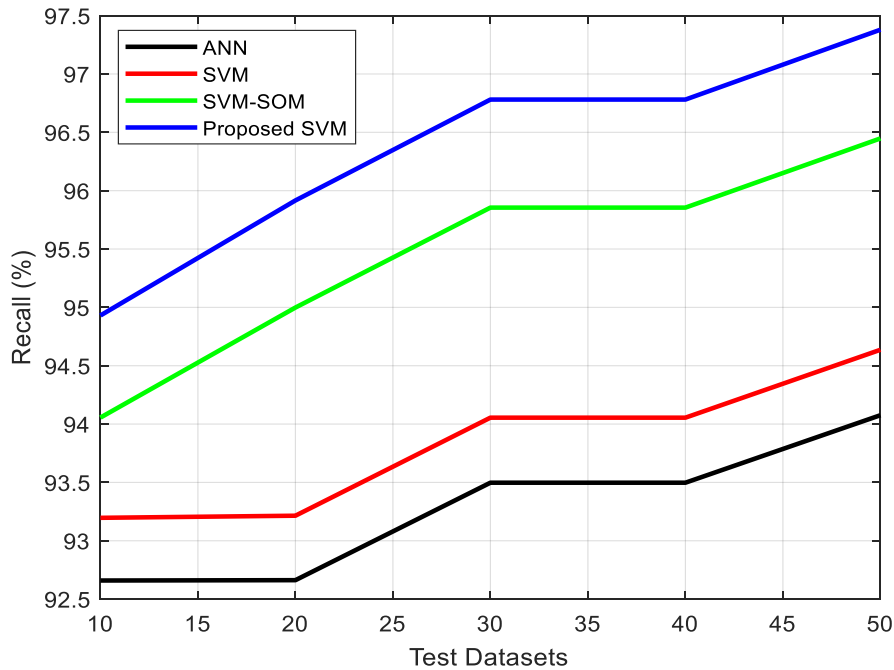


Fig. 5. Recall

Fig. 4 shows the results of precision in detecting the attacks with the conventional NIDS models. The results show that the proposed NIDS has a higher rate of precision than the other methods.

Fig. 5 shows the results of recall in detecting the attacks with the conventional NIDS models. The results show that the proposed NIDS has a higher rate of recall than the other methods.

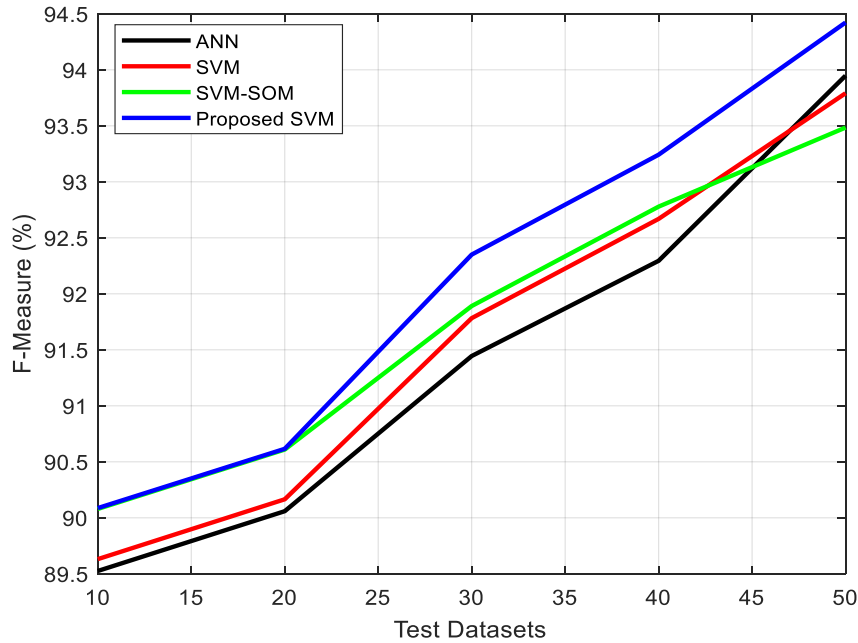


Fig. 6. F-measure

Fig. 6 shows the results of f-measure in detecting the attacks with the conventional NIDS models. The results show that the proposed NIDS has a higher rate of f-measure than the other methods.

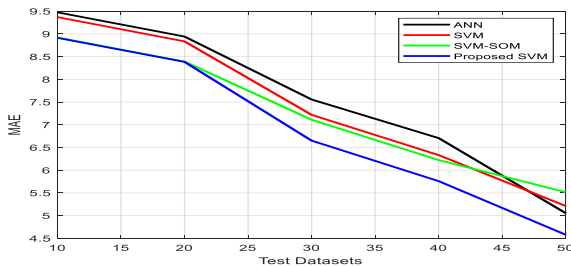


Fig. 7. MAE

Fig. 7 shows the results of mean absolute error (MAE) in detecting the attacks with the conventional NIDS models. The results show that the proposed NIDS has a reduced error than the other methods. From the results of simulation, it is seen that the proposed SVM has reduced error rate than the existing methods.

## V. CONCLUSION

In this paper, we employ a machine learning-based IDS to protect the SDN-IoT networks from DDoS attacks. The framework involves feature selection-based classification allows the IDS to offer better defence against DDoS attacks. The resilience of the model to DDoS attacks of varied severities is tested via simulation. The simulation results show that the proposed approach is superior to other options for identifying and categorizing network intrusions. In future, the attacks on large scale network can be mitigated using multi-

SDN controllers which reduce single point of failure and this cannot pose a serious computational burden on the networks.

## REFERENCES

- [1] J. Bhayo, R. Jafaq, S. Hameed and S. A. Shah, "A time-efficient approach toward DDoS attack detection in IoT network using SDN," *IEEE Internet of Things Journal*, 9(5), 3612-3630, 2021.
- [2] K. M. S. Azad, N. Hossain, Md. J. Islam, A. Rahman, S. Kabir, "Preventive determination and avoidance of ddos attack with sdn over the iot networks," *IEEE International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)* (pp. 1-6). 2021.
- [3] M. Aslam et al., "Adaptive Machine Learning Based Distributed Denial-of-Services Attacks Detection and Mitigation System for SDN-Enabled IoT," *Sensors*, 22(7), 2697, 2022.
- [4] H. Aldabbas, and R. amin, "A novel mechanism to handle address spoofing attacks in SDN based IoT," *Cluster Computing*, 24(4), 3011-3026, 2021.
- [5] A. Wani and R. Sathiyaa, "SDN-based intrusion detection system for IoT using deep learning classifier (IDSIoT-SDL)," *CAAI Transactions on Intelligence Technology*, 6(3), 281-290, 2021.
- [6] D. Javeed, T. Gao, M. T. Khan and I Ahmad, "A hybrid deep learning-driven SDN enabled mechanism for secure communication in Internet of Things (IoT)," *Sensors*, 21(14), 4884, 2021.
- [7] D. Javeed, T. Gao, M. T. Khan and I Ahmad, "SDN-enabled hybrid DL-driven framework for the detection of emerging cyber threats in IoT," *Electronics*, 10(8), 918, 2021.
- [8] A. Shoeb and T. Chithralekha, "Resource management of switches and Controller during saturation time to avoid DDoS in SDN," *IEEE International Conference on Engineering and Technology (ICETECH)* (pp. 152-157). 2016.
- [9] R. T. Kokila, S. T. Selvi and K. Govindarajan, "DDoS detection and analysis in SDN-based environment using support vector machine classifier," *IEEE sixth international conference on advanced computing (ICoAC)* (pp. 205-210). 2014.
- [10] P. Xiao, Z. Li, H. Qi, W. Qu, and H. Yu, "An efficient DDoS detection with bloom filter in SDN," *IEEE Trustcom/BigDataSE/ISPA* (pp. 1-6), 2016.
- [11] S. Lim, J. Ha, H. Kim, Y. Kim, and S. Yang, "A SDN-oriented DDoS blocking scheme for botnet-based attacks," *IEEE Sixth International*

- Conference on Ubiquitous and Future Networks (ICUFN) (pp. 63-68), 2014.
- [12] T. V. Phan, N. K. Bao, and M. Park, "A novel hybrid flow-based handler with DDoS attacks in software-defined networking", Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCCom/IoP/SmartWorld) (pp. 350-357), 2016.
- [13] T. Chin, X. Mountrouidou, X. Li, and K. Xiong, "An SDN-supported collaborative approach for DDoS flooding detection and containment," In MILCOM 2015-2015 IEEE Military Communications Conference (pp. 659-664), 2015.
- [14] S. Hameed, and H. A. Khan, "Leveraging SDN for collaborative DDoS mitigation," IEEE International Conference on Networked Systems (NetSys) (pp. 1-6), 2017.
- [15] R. Macedo, R. D. Castro, A. Santos, Y. Ghamri-Doudane, and M. Nogueira, "Self-organized SDN controller cluster conformations against DDoS attacks effects," IEEE Global Communications Conference (GLOBECOM) (pp. 1-6), 2016.
- [16] R. Sahay, and G. Blanc, "ArOMA: An SDN based autonomic DDoS mitigation framework," *computers & security*, 70, 482-499, 2017.

# Customer Sentiment Analysis in Hotel Reviews Through Natural Language Processing Techniques

Soumaya Ounacer<sup>1</sup>, Driss Mhamdi<sup>2</sup>, Soufiane Ardchir<sup>3</sup>, Abderrahmane Daif<sup>4</sup>, Mohamed Azzouzi<sup>5</sup>

Faculty of Sciences-Department of Information and Modelisation Technologies,

Hassan II University, Ben M'sik, Casablanca, Morocco<sup>1, 2, 4, 5</sup>

National School of Business and Management, Casablanca, Morocco<sup>3</sup>

**Abstract**—Customer reviews of products and services play a key role in the customers' decision to buy a product or use a service. Customers' preferences and choices are influenced by the opinions of others online; on blogs or social networks. New customers are faced with many views on the web, but they can't make the right decision. Hence, the need for sentiment analysis is to clarify whether opinions are positive, negative or neutral. This paper suggests using the Aspect-Based Sentiment Analysis approach on reviews extracted from tourism websites such as TripAdvisor and Booking. This approach is based on two main steps namely aspect extraction and sentiment classification related to each aspect. For aspect extraction, an approach based on topic modeling is proposed using the semi-supervised CorEx (Correlation Explanation) method for labeling word sequences into entities. As for sentiment classification, various supervised machine learning techniques are used to associate a sentiment (positive, negative or neutral) to a given aspect expression. Experiments on opinion corpora have shown very encouraging performances.

**Keywords**—Topic modeling; aspect-based sentiments analysis; aspect extraction; sentiment classification; machine learning

## I. INTRODUCTION

In the last few years, the use of the internet and online interactions has grown tremendously. A significant quantity of data are generated daily via social media, forums, chats, and other sources that is primarily displayed as natural language text[1]. The way internet users behave online has also changed how the internet works. For instance, rather than being merely content consumers, internet users are becoming content creators[2]. One significant piece of information that is produced daily within the wide range of content produced by internet users is opinions[2].

Internet users have the ability to criticize or popularize a service or a product with a simple comment or review on the internet and in different fields[3]. Numerous enterprises and businesses have taken advantage of this pertinent data to offer the greatest services or goods for their clients. Among these areas, tourism which is a continuously developing industry and an important key industry for many regions and countries[4]. The opinions and reviews of tourists who visit touristic places every year are shared on various sites such as TripAdvisor, Booking and Yelp...etc[5][6]. Internet users do not have the ability to read, understand and summarize the large number of reviews available for a specific hotel. It is challenging for a simple user to make use of the information at hand to choose a comfortable hotel for his/her trip. The principle on which this

work is based is to carry out an analysis of customers' opinions on hotels located in Marrakech in order to allow them to improve their services and focus more on the main obstacles that have an impact on the attractiveness of these hotels. In this article, a study and application of Aspect-Based Sentiment Analysis are carried out in the hotel and tourism industry. Specifically, opinions will be analyzed so as to determine the sentiment that is expressed towards certain characteristics of the hotel and the service delivered by its employees. The main goal is to produce results whose conclusions can provide directions that lead to improve the performance in sentiment analysis[7]. To achieve this goal, several objectives will be accomplished. The first objective is to use the various preprocessing steps available for text preprocessing. The second objective is using existing libraries like, TextBlob or Vader to label the dataset. The third aim is to use and compare multiple classification methods to classify the views so as to correct aspects and sentiments, i.e., classification of online comments into polarity (negative, positive, and neutral), and finally apply an Aspect-Based Sentiment Analysis on the product (hotel) features identified. In order to achieve these objectives through a clear and logical progression, this work will be presented according to the following structure. Section II will be devoted to the different related works linked to the Aspect-Based Sentiment Analysis. As for Section III, it will expose a background of Topic Modeling and Machine Learning model. The construction of the DataFrame, the methodology and the experimental results will be presented in Section IV. A summary of the experiment's results will be shown in Section V. Last but not least, Section VI will be devoted to the conclusion and future work.

## II. RELATED WORKS

For aspect-based opinion classification, aspect extraction is a crucial task. The vast bulk of extractions techniques have recently been put forth for the tourism industry. These methods have employed a variety of mechanisms and techniques to extract crucial information from tourism reviews. These methods can be split into three primary groups: methods based on rules, seeds, and topic models. There are several works in the field of hotels and tourism that concern Aspect-Based Sentiment Analysis which will be described as the following:

"Pekar et al." [8] utilized TermExtractor to divide hotel reviews into terms. The terms were then trained in a lexicon. Finally, they manually extracted from the term lexicon the six most obvious characteristics (single nouns and multi-word

nouns). This proposed method is based on rules that allow to extract aspects from hotel reviews using aspect appearance on every review.

Similar preprocessing steps were used by "Muangon et al." [9] and were supported by LexToPus. These steps are used to categorize all hotel reviews into features. These characteristics include polar words as well as aspects. They extracted all of the top-rated aspects using a prioritized method.

"Marrese taylor et al." [10] have suggested an algorithm with the goal of extracting aspects. Aspects from restaurant reviews can be extracted thanks to this algorithm. The authors converted the reviews into sentences and then used Part-Of-Speech tagger to extract nouns from the sentences.

A different method for aspect extraction was proposed by "Hai et al." [11]. According to two criteria—domain specific and domain independent—the authors extracted aspects. They created a list of candidate aspects by first using syntactic dependency rules. Then, they determined the intrinsic domain relevance score (IDR) and extrinsic domain relevance score (EDR) for each specific domain and independent domain of each extracted candidate feature, respectively. And at the end, these candidate features are extracted from the list of candidates that have low IDR score and high EDR score.

An algorithm based on a bootstrapping approach, which has been proposed by "Wang et al." [12], extracts the main aspects of the review. In this algorithm, each sentence was initially given an aspect based on the maximum of overlap between its words and the aspect. Then, to examine the relationship between the allocated aspect and the sentence words, they determine the basic dependencies between them. Finally, sentence words that have a strong relationship with the assigned aspect are added to the list of aspect keywords and are considered to be aspects.

BESAHOT, which is a system that has been presented by "Walter Kasper et al." [13], performs analyzed comment processing for text segmentation, statistical polarity detection of text segments, and extraction of linguistic information from review topics and their aspects. It is a quality control support system for hoteliers that provide them with complete overviews and summaries of their hotel and how it is rated and commented by users on the web.

### III. ASPECT-BASED SENTIMENT ANALYSIS

An essential task in the field of Sentiment Analysis is Aspect-Based Sentiment Analysis (ABSA) [14]. It involves assigning a polarity (positive, negative or neutral) to each aspect evoked in an opinion sentence. Aspect extraction and aspect-level sentiment analysis are often the two main tasks used to accomplish this.

Although traditional Sentiment Analysis is done using document and sentence level Sentiment Analysis techniques, the current trend is to move to a deeper level which is presented as ABSA (features). This latter [15] performs a deeper and a better analysis, as it directly examines the opinion itself. This domain is a deeper end in Natural Language Processing (NLP) [16] where it presents a richer problem for researchers.

One of the main features of NLP is Topic Modeling. Topic Modeling can be applied to any form of text: emails, tickets, feedbacks, etc. in order to have a global vision of customers' concerns.

#### A. Topic Modeling

Natural Language Processing (NLP) [17] includes Topic Modeling, which is used to train Machine Learning models. It entails identifying from a document or corpus of data the words of themes that are associated with a particular topic.

Topic Modeling is an unsupervised Machine Learning approach to discover topics in various text documents. It can find patterns of words and phrases and automatically cluster groups of words and associated phrases that best represent the whole [18]. It also provides a useful view of a large corpus in terms of the relationships between them and individual documents.

The figure above (Fig. 1) represents the ordinary workflow of the thematic modeling process. A set of text documents is introduced into the black box of the thematic modeling algorithm and the following results are obtained:

List of Topics: Topics are the key themes representing the entire collection of documents. Each topic consists of several words that occur at the same time. A word can belong to more than one topic because a word can have a different meaning in a different context.

Topic Definition: A topic is represented by the weighted frequency of words. Each topic can be interpreted as a theme.

Topic Distribution of Document: Each document is represented as a topic distribution where the weight of a topic defines the part of the document covered by that topic. In a way, it provides a "soft grouping" of the document.

Topic modeling is a method for selecting a set of topics from a group of documents that best summarizes the information in the group. To create topic models, numerous techniques are employed. One of the areas of interest is in: LDA, LSA, NMF, and Corex which will be discussed later on in this section.

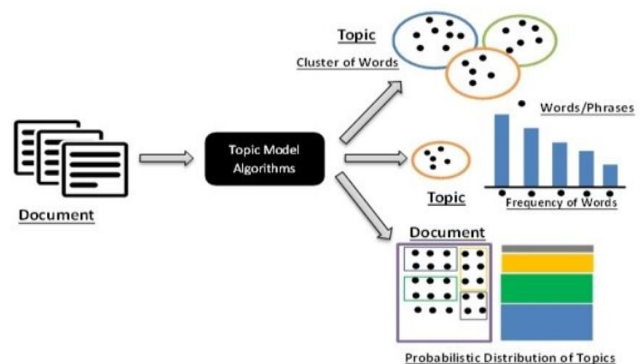


Fig. 1. Topic models process

1) LDA: Latent Dirichlet Allocation, [19] is a powerful learning algorithm for automatically and jointly classifying words in documents in mixtures of contexts. It has been successfully applied to model changes in scientific domains

over time. LDA is a probabilistic generative model. Based on the assumption that the order of documents in the collection and the order of words in a text are indifferent, LDA defines finite mixture models on sets of underlying topics to generate the collection. Each topic is being modeled as an infinite mixture on probabilities of the underlying topics. In an iterative procedure, these probabilities are computed several times, until the algorithm converges.

Advantages: Among the advantages of using the LDA method, the following are worth mentioning:

- LDA is easy to implement, understand and use.
- It maximizes inter-class scattering.
- It reduces intra-class scattering.

Disadvantages: Despite these advantages, a set of negative points still exist such as:

- LDA is costly in computation time.
- It is also costly in memory space.
- It renders poor results when the number of training images is large.
- It is hard to know when LDA is working. Metrics like perplexity are acceptable to check if learning is working, but there is a very poor indicators of overall model quality. For example, you could have a model with very low perplexity, but whose topics are not very informative.
- The topics are predicated on the multinomial distribution, while the words are predicated on a different multinomial distribution formed specifically for this topic. The structure may not be properly adjusted if the real structure is more complex than a multinomial distribution or if the data needed to construct the structure are insufficient.
- The user specifies the total number of subjects in the dataset (or bases it on a certain distribution using sampling), which is subjective and may not always reflect the true distribution of subjects.

2) *LSA*: Latent Semantic Analysis [20], or *LSI* (Latent Semantic Index), employs a bag-of-words (BoW) model, creating a term-document matrix (occurrence of terms in a document) [21]. Terms are represented in rows, and documents are represented in columns. By applying singular value decomposition to the term-document matrix, LSA can identify latent subjects. It is typically applied as a technique for noise or dimension reduction. In this method, document analysis is done by machines using Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF is a metric that quantifies the significance of a word for a corpus of documents.

Advantages:

- Easy to understand and implement.

- In comparison to the vector space model, it shows improved outcomes.
- Only involves the decomposition of document term matrix which makes it faster than other available algorithms.

Disadvantages:

- In general, LSI is very slow on large corpora and not very accurate compared to LDA.
- The dimension of the latent subject depends on the rank of the matrix.
- The decomposed LSA matrix is extremely dense, making it challenging to index the individual dimension.
- Polysemy cannot be captured by LSA (multiple meanings of a word).
- It provides less accuracy than LDA.

3) *NMF*: Since the non-negative matrix factorization [22] is an unsupervised method, the subjects on which the model will be trained are not labeled. NMF factors or decomposes high-dimensional vectors into a representation that has a lower dimension. Since the coefficients of these lower-dimensional vectors are nonnegative, they are likewise nonnegative vectors. Consider the general scenario where there is an input matrix  $V$  with the form  $m \times n$ . This approach divides  $V$  into two matrices,  $W$  and  $H$ , whose dimensions are  $m \times k$  and  $n \times k$ , respectively. In this case,  $V$  stands for the term document matrix,  $H$  stands for an embedded word in each row, and  $W$  stands for the weight of each word found in each sentence.

Advantages:

- NMF can handle missing values naturally and this property leads to a new method to determine the rank hyper parameter.

Disadvantages:

- NMF cannot be applied to several real-world issues where the domain limited knowledge of experts is available
- It sometimes provides semantically incorrect results

4) *CorEx*: Contrary to LDA and NMF, the semi-supervised topic model Correlation Explanation [23] allows to give the model "anchor words" which exemplified potential topics that the model might be looking for. CorEx also allows to provide the model with a confidence score for the anchors. If this choice is less certain, the model may forgo the target recommendations if they don't sufficiently match the data. This new capability is strong in guiding a thematic model with the chosen domain expertise. CorEx provides a flexible framework for learning topics that are maximally informative about a text corpus. The CorEx topic model makes few assumptions about the LDA structure and flexibly incorporates domain knowledge through user-specified



"anchor words". With anchor words, one can guide the topic model to topics of substantial interest, interact with the topics, and refine them in ways not possible with traditional topic models.

Advantages: CorEx competes with LDA in terms of producing semantically consistent topics that aid in document classification. By citing above some advantages of this model [23]:

- The CorEx's modeling algorithm is rapid.
- It searches for topics that are "maximally informative" about a set of documents rather than assuming a specific model of data generation.
- Word-level domain knowledge can be flexibly integrated into the CorEx thematic model.
- It consistently creates document clusters with higher homogeneity than LDA in terms of clustering.
- The CorEx anchor steers the thematic model toward topics that don't naturally arise and frequently results in topics that are more coherent and predictable.

Disadvantages: Despite the number of advantages of the CorEx thematic model over LDA, there are some drawbacks[23]:

- The sparse implementation necessitates that every word appears in just one topic. It is not a matter of fundamental theoretical limitations, but rather of computer efficiency.
- CorEx relies on binary accounting data for its parcel-level optimization rather than the usual accounting data that are input into LDA and other theoretical models.

Despite binary number limitations, CorEx nonetheless discovers a reliable and competitive structure in the data.

### B. Machine Learning Model

The determination of the direction of the opinions in a text divided into two or more classes on certain features is known as classification of opinions by aspect. The classification of opinions has been done into several categories, such as binary, ternary, etc.

Typically, the classification task is defined as the task of predicting the label that is to say, assigning each given object to a group based on a classification rule. The primary goal at work entails classifying aspect opinions, thus training a classifier to predict the label for each input text is needed. There are three kinds of polarities (positive, negative, or neutral). In this section, the most common employed algorithm shall be outlined. [24].

1) *Logistic regression (LR)*: This is an analytical technique key in the social and scientific sciences [25]. Logistic regression, which also closely resembles neural networks, is the standard supervised Machine Learning approach for classification in natural language processing. A logistic function is used in logistic regression to create discrete dependent variables from a series of data points.

2) *Support vector machines (SVM)* [26]: This can be applied to both regression and classification tasks. SVM methods aim to partition linearly separable data into two classes with the maximum distance between them. In the high-dimensional space, SVM identifies an ideal hyperplane that separates the input data with the greatest possible margin between it and the point(s) that are closest to it. The points for which the margin is reached are called support vectors. A kernel function can be used to map the data into a higher dimensional space in order to make them linearly separable if the input data are not linearly separable. The polynomial kernel, Gaussian radial basis function, and sigmoid kernel are the three most widely used nonlinear kernels.

3) *K-nearest neighbour (K-NN)*: This is one of the simplest Machine Learning algorithms used for classification and regression problems[27]. The information is subsequently allocated to the class with the closest neighbor based on the nearest measures.

4) *Naïve bayes (NB)* [28]: This Machine Learning algorithm can be used to divide objects into two or more classes, such as text documents. It is founded on the Bayes theorem, which uses conditional probabilities as its foundation.

5) *Decision tree (DT)*: This is part of the supervised algorithms in the field of Machine Learning [29]. Their principle is to divide learning data into groups whose content becomes increasingly homogeneous until pure data are obtained (belonging to the same class) or a maximum number of partitions is reached. The resulting model is a tree composed of several decision rules and is easily interpretable. As with any supervised learning method, decision trees make use of examples. Building a decision tree by category is necessary if one has to categorize the documents. In order to determine to what extent a category a new document belongs to, the Decision Tree will be used for each category in which the classified document is submitted. Each tree responds with yes or no.

6) *Random forest (RF)* [30]: This is a prediction method that Ho developed in 1995. In 2001, scientists Leo Breiman and Adele Cutler formally proposed the algorithm. It is made up of various decision trees that each focus on a different aspect of the problem independently. Multiple decision trees are produced by this classifier using a subset of the training data that is randomly chosen. The final class of test objects is then decided by aggregating the votes from various decision trees.

7) *ExtraTrees (ET)*: Extremely Randomized Trees [31] is an ensemble-supervised Machine Learning technique that makes use of decision trees. It builds multiple trees and divides the nodes using random subsets of features, but the sampling for each tree is without replacement. As such, the most important and unique feature of the algorithm is the random selection of a splitting value for a feature, which makes the trees diverse and uncorrelated.

8) *AdaBoost (AB)* [32]: This is a widely used boosting algorithm. It builds a majority vote iteratively. Over the iterations, it maintains a weight distribution on the training examples so that poorly classified examples see their weight increase and well classified examples see their weight decrease. At each iteration, the weak learning algorithm is trained with the set of weighted examples and the resulting classifier is added to the majority vote.

9) *GradientBoost (GB)*: A meta estimator that fits a series of weak learners is gradient boosting [33]. It is a powerful Machine Learning algorithm used to solve regression and classification problems. It creates a prediction model in the form of a collection of weak prediction models, typically decision trees. It builds the model incrementally, much like other boosting techniques do, and generalizes them by enabling the optimization of an arbitrarily differentiable loss function.

#### IV. BUILDING THE DATAFRAME

The workflow and the subtasks for each phase are shown in Fig. 2. The learning phase and the testing phase are the two steps that make up this workflow.

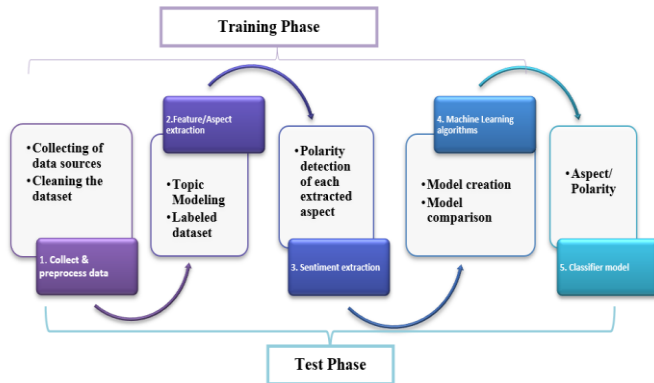


Fig. 2. Aspect-based sentiment analysis workflow

##### A. Training Phase

Fig. 3 gives an overview of this process namely the collection, the data sources and the pre-processing of the datasets. The main tasks of each step are described in the paragraphs that follow.

1) *Data sources*: User opinions are the main criterion for enhancing the quality of the services provided and improving the value of the products delivered. These opinions can be found in different data sources namely review sites, blogs and micro-blogs.

a) *Review sites*: Opinions have the role of decision makers for any user during the purchase phase. User generated reviews of products and services are widely available on the

internet. Sentiment ratings or texts use reviewer data collected from websites such as TripAdvisor and Booking (hotel reviews). These sites host millions of visitors' hotel reviews.

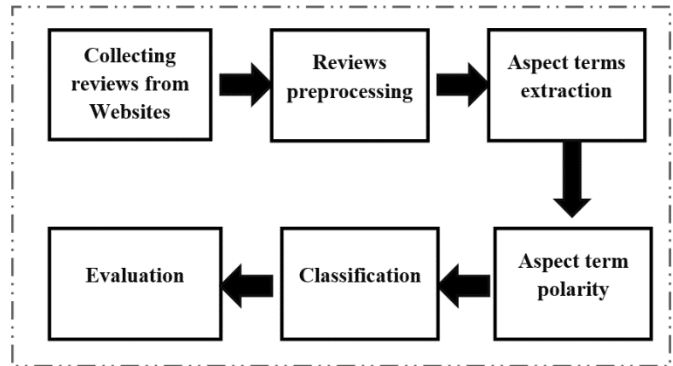


Fig. 3. Sentiment analysis process at aspect level

b) *Blogs and micro-blogs*: Blogs and micro-blogs are among the most popular communication tools of internet users. Millions of messages are posted every day on well-known microblogging platforms including Twitter, Tumblr, and Facebook. Sometimes Twitter messages express opinions that are used as a source of data to classify sentiments.

2) *Data collection*: The data acquisition or collection phase consists of obtaining the corpus to be analyzed. The "web scraping" method [34] is used to collect the reviews, since the goal is to collect reviews from various hotels in Marrakech. As shown in Fig. 4, each entry in this dataset is structured as follows:

- *Hotel\_name*: designates the name of the establishment (Hotel)
- *Title\_review*: refers to the title written by the client to give a general summary
- *Reviews\_hotel*: contains reviews (text), written in English
- *Rating\_date*: indicates the date of publication of a journal
- *Score\_rating*: the evaluation given by each customer between 10 and 50

The reviews for 10 different hotels in the city of Marrakech are obtained from two websites (booking and TripAdvisor). The dataset consists of 21619 reviews in English, but only 14356 reviews are used for this study. Table I represents a summary of the dataset used:

TABLE I. SUMMARY OF THE DATASET

Domain	Numbers of reviewers	Words average
Hotels	14356	7.06

	Hotel_Name	Title_review	reviews_hotel	Score_rating	Rating_Date
0	Radisson Blu Hotel, Marrakech Carre Eden	Such a nice experience! My best hotel in marra...	A peaceful hotel that I completely loved & enj...	50	29 April 2021
1	Radisson Blu Hotel, Marrakech Carre Eden	Good experience and good service, nice staff!	I appreciated the service and the warm welcom...	50	18 April 2021
2	Radisson Blu Hotel, Marrakech Carre Eden	Not as expected	Needs more adjustments until doesn't seem that y...	30	10 April 2021
3	Radisson Blu Hotel, Marrakech Carre Eden	Great Room Service	Have to say my start with the Radisson was iff...	40	30 March 2021
4	Radisson Blu Hotel, Marrakech Carre Eden	Refund Refused despite the Pandemic	I was an expat living in Marrakech. Four of us...	10	19 November 2020

Fig. 4. Example of datasets

3) *Review pre-processing*: The pre-processing procedure followed in this work aims to clean up the notices and to make them as close as possible to a formal language. First, the notices were filtered considering only those written in English. Because a corpus of different languages is a corpus that contains noise. To do this, a Python library called Langdetect is used and then proceeded to a pre-processing that follows the following steps:

Split the text into several rows: As represented in Fig. 5, this task consists of splitting the text contained in each cell of the "Reviews\_hotel" column into several rows by the '.' delimiter, for example:

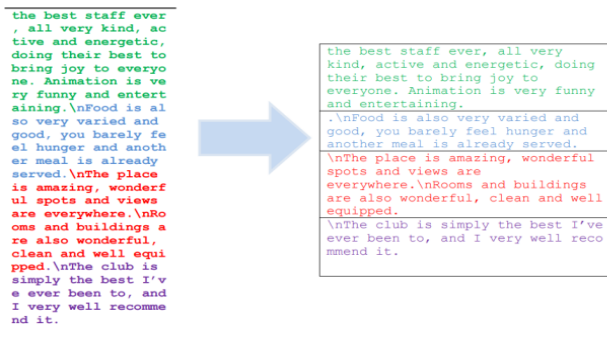


Fig. 5. Example of paragraph splitting into multiple lines

Noise cleaning - spacing, special characters, lowercasing: In a text you can find various characters such as numbers, free white spaces, all kinds of punctuation and some terms are put in random capital letters. This form of noise cleaning takes care of the spacing and special characters. Turning all words into lower case is also a very common pre-processing step. Next, all punctuation and special characters will be removed since they serve no purpose once analyzing the data begins.

Eliminate emoji: One cannot ignore the content of the notices full of emoticons, symbols and pictographs as well as a set of flags. So, this task concerns the elimination of these characters.

Tokenization / eliminating words below three letters: First, separating the corpus into a vocabulary of single terms is essential, which is called tokenization. Individual terms and overwrite all words below three letters can be tokenized.

Delete Stop Words: Some words in English, while necessary, do not contribute much to the meaning of a sentence. These words, such as "when", "had" or "before", are called stop words and should be filtered out.

Stemming and lemmatization of the text: The process of turning a word into its Racine form is known as racinization.

Rooting can create non-real words. Lemmatization, as opposed to racinization, aims to obtain the canonical (grammatically correct) word forms, or lemmas. In terms of calculation, lemmatization is much more difficult and computationally expensive than racinization. In actual practice, the two methods have little impact on the performance of text classification.

4) *Aspect terms extraction*: To accomplish the Machine Learning task, the aspect or category detection of each review must be first tackled. This can be done according to different tools described in Section III. Topic Modeling has been tested using LDA techniques as well as CorEx and NMF.

For this experimentation, 9 Aspects are predefined: Rooms/Cleanliness, Location, Staff, Food/Restaurant, Experience/Value, Price/Quality, Service, Amenities/Activities, and Hotel/Property.

5) *Topic modeling*: In this section, the steps as well as the result of the three topic modeling methods have been presented, namely, LDA, NMF and CorEx. The steps followed using LDA are: corpus vectorization which allows to create the term matrix document using Counvectorizer which is an excellent tool provided by Scikit-learn library in Python. It is used to transform a given text into a vector based on the frequency (number) of each word that occurs in the whole text. Then, the LDA model will be built, to evaluate this performance with perplexity and log probability. Gridsearch was used to choose the "right" number of topics for the LDA model, then two hyper parameters (learning decay and number of topics) were tested. And finally, the labeled topics joined the original text.

In Non-negative Matrix Factorization (NMF), the first step is to convert the document into a term-document matrix which is a collection of all the words in the given document using the TfidfVectorizer. Then, build the NMF model with Scikit-learn and view the original topics. And as far as CorEx is concerned, the first step is the corpus vectorization which converts the document into a document-Terms-matrix using the TfidfVectorizer by creating a vocabulary containing the topics. The second step is the creation of the model, starting with the identification of the Anchors (anchored words) and then creating the model that allows to generate the set of topics. In each of these models, topics related to four distinct themes for LDA and NMF can clearly be seen. But it is also clear that these topics contain words that can apply to multiple contexts and cause problems in certain circumstances. The following table (Table II) illustrates the comparison between these techniques:

TABLE II. COMPARISON OF THE THREE EXTRACTION METHODS

	LDA	NMF	CorEx
<b>Topic 0</b>	'room','leave','shower','bathroom','table','change','towel','work','wasnt','tv','dirty','open','service','expect','door','drink','provide','toilet','available','reception'	'room','clean','comfortable','spacious','bed','bathroom','beautiful','shower','big','small','towel','need','view','large','size','balcony','wifi','daily','work','floor'	room, clean, bed, comfortable, bath room, shower, towel, spacious, table, door, balcony, room clean, bedroom, room spacious, room pool, petal, clean room, hotel clean, size, toilet
<b>Topic 1</b>	'staff','friendly','helpful','hotel','spa','nice','reception','ok','massage','ask','tour','manager','extremely','speak','french','stay','english','owner','polite'	'staff','friendly','really','attentive','polite','welcome','reception','amaze','extremely','service','professional','amazing','member','animation','nice','make','team','kind','really'	staff, friendly, helpful, team, animation team, manager, waiter, professional, staff friendly, receptionist, reception staff, hotel staff, attentive, friendly helpful, staff helpful, member, polite, lifeguard, restaurant staff, helpful staff, staff polite
<b>Topic 2</b>	'room','clean','pool','nice','hotel','bed','view','lovely','comfortable','beautiful','small','spacious','sun','garden','large','great','big','wifi','terrace','ground'	'great','value','location','experience','time','service','breakfast','atmosphere','staff','family','view','overall','team','animation','trip','kid','money','visit','people','spa'	food, restaurant, breakfast, drink, dinner, meal, fresh, lunch, menu, buffet, delicious, cook, coffee, fruit, snack, omelette, salad, eat, ate, tea, bread, juice, order, mint
<b>Topic 3</b>	'team','staff','make','animation','work','bar','entertainment','really','help','amaze','brilliant','hard','special','aqua','attentive','great','time','especially','best','kid'	'hotel','best','beautiful','amazing','lovely','time','book','nice','fantastic','like','city','locate','boutique','wonderful','shuttle','medina','visit','really','experience','return'	hotel, stay, back, recommend, experience return, enjoy, come back, good, highly recommend, recommend hotel, stay hotel, highly enjoyed, place stay, beautiful hotel, nice hotel, boutique hotel, lovely hotel

While topic models can be rapidly run, they are not necessarily as accurate in their classification decisions as the more complicated supervised learning models, and occasionally their outputs can even be outright false. Semi supervised topic modeling will be utilized to determine the main topics of these documents in order to prevent ambiguities between topics. This most recent development gave a middle ground between supervised classification modeling and unsupervised topic modeling.

From the comparison table, CorEx provides more specification of aspects than the others, CorEx is chosen as the best, as the grouping of each aspect seems better. Fig. 6 shows what the dataset looks like at this point.

	review_text_clean	Score_rating	dominant_topic	Categorie_Aspect
0	nice experience hotel amaze view	50	4.0	Experience/Value
1	netflix amaze	50	7.0	Amenities/Activities
2	peaceful hotel completely love enjoyed beginning	50	4.0	Experience/Value
3	staff professional welcoming	50	2.0	Staff
4	shoutout rachid lansari manager team	50	2.0	Staff
...	...	...	...	...
14351	hotel close everywhere	30	1.0	Location
14352	recommend little cafe door delicious fresh juice	30	4.0	Experience/Value
14353	pool lovely	30	7.0	Amenities/Activities
14354	food	30	3.0	Food/Restaurant
14355	price pay overall happy	30	5.0	Price/Quality

14356 rows x 4 columns

Fig. 6. Datasets after aspect extraction

6) *Annotation of journals:* For review annotation, reviews were labeled using two tools VADER (Valence aware Dictionary and Sentiment Reasoner) and TextBlob. There are three types of sentiments in this dataset: positive, negative, and neutral. To pursue the supervised learning approach, the type of sentiment (polarity) of each review should be known.

VADER has been chosen since it provides a better classification and more negative feelings than the other (Fig. 7).

	review_text_clean	Score_rating	dominant_topic	Categorie_Aspect	scores compound	Sentiment_Vader
0	nice experience hotel amaze view	50	4.0	Experience/Value	{'neg': 0.0, 'neu': 0.323, 'pos': 0.677, 'compound...}	0.7430 Positive
1	netflix amaze	50	7.0	Amenities/Activities	{'neg': 0.0, 'neu': 0.222, 'pos': 0.778, 'compound...}	0.5423 Positive
2	peaceful hotel completely love enjoyed beginning	50	4.0	Experience/Value	{'neg': 0.0, 'neu': 0.21, 'pos': 0.79, 'compound...}	0.9056 Positive
3	staff professional welcoming	50	2.0	Staff	{'neg': 0.0, 'neu': 0.408, 'pos': 0.592, 'compound...}	0.4404 Neutral
4	shoutout rachid lansari manager team	50	2.0	Staff	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...}	0.0000 Neutral

Fig. 7. Datasets after polarity detection

7) *Multi-target classification (aspect/sentiment):* In order to use machine learning algorithms in the textual data, there is a need to represent the text in the document as a vector of fixed size and this in order to plunge the data in a metric space. Among the vectorization techniques the TF-IDF and CountVectorizer are two ways to convert text into numbers.

a) *Count vectorizer:* Count Vectorizer offers a straightforward method for tokenizing a group of text documents, creating a vocabulary of recognized words, and encoding new documents using that vocabulary.

b) *TF-IDF vectorizer:* TF-IDF, which stands for Term Frequency - Inverse Document Frequency, is a statistic which is based on a word's frequency in the corpus. It also gives a numerical representation of a word's importance for statistical analysis.

### B. Testing and Evaluation Phase

In this experiment, the most commonly used classifiers in the sentiment analysis literature are applied. "Documents x Terms" vectorization method will be evaluated using nine supervised classification models: Bayesian Naive, SVM, Logistic Regression, K-nearest Neighbor, Decision Trees, Random Forests, Extratrees, Adaboost and Gardient Boost. The performance of the selected models will be compared using their Accuracy, Precision, Recall and F1-scores to determine the best decision model.

1) *Description datasets*: The sentiment analysis as well as the aspect analysis were performed on a dataset that contains 14356 English dialect reviews from TripAdvisor and Booking websites and labeled as follows: 4337 positive texts, 397 negative texts and 8647 neutral, still 2822 texts labeled with the aspect Room, 386 with Service, 1952 with Food/Restaurant, 2118 with Staff, 1352 labeled with Location, 2833 with Experience/Value, as well as 992 are labeled with the aspect Amenities/Activities and 420 reviews are with Price/Quality.

2) *Performance measure*: The choice of classifier for the current data is based on the performance measures [35]. The evaluation of the optimal solution in classification training can be defined based on the confusion matrix. From the given confusion matrix one can determine, the number of positive and negative that are correctly classified. Meanwhile, the number of negative and positive cases are misclassified respectively. The performance measure of the various classifiers is evaluated using the accuracy, precision, recall and F1-scores.

The Accuracy metric: indicates the percentage of correct predictions. It refers to the ratio of the number of correct predictions to the total number of input samples or observations, which is shown in eq. (1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision: is the number of correct positive results divided by the number of positive results predicted by the classifier. The result is a value between 0.0 for no precision and 1.0 for total or perfect precision.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall: In order to complete the accuracy, the recall is also calculated, which is the fraction of true positives to real positives, which is shown in eq. (3), i.e., the proportion of positives that were correctly identified.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-Score: The calculated average harmony of precision and recall is used to assess how well these two metrics—rappel and precision—compromise. This unique score ranges from 0 to 1, with 0 being the worst possible outcome and 1 being the best possible outcome. It can be calculated as follows.

$$F1 = \frac{2*(Precision*Recall)}{Precision+Recall} \quad (4)$$

3) *Results*: The learning phase will be followed by the testing phase in order to evaluate the classifier. For performance validation, 80/20% rules is used to check the model, the corpus is divided into two parts, 80% for the training phase and 20% for the testing phase. Several tests, whose results are presented in the following tables, were made:

#### a) CountVectorizer embeddings classification

- Aspect Classification

TABLE III. RESULTS OBTAINED FROM ALL CLASSIFIERS IN COUNTVECTORIZER WEIGHTING METHOD FOR ASPECT CLASSIFICATION

Metrics \ Classifiers	Accuracy %	Recall%	Precision %	F1-Score %
LR	83.25	73.01	75.56	73.42
RF	<b>84.26</b>	<b>74</b>	<b>77.14</b>	<b>75</b>
NB	76	55.78	67.68	56.29
DT	48.70	28.12	43.75	26.53
KNN	71.59	63.71	67.22	64.95
SVM	83.51	70.91	86.84	72.88
ET	81.67	71.41	74.66	72.49
AB	69.83	59.33	65	60.44
GB	76.13	61.07	69	62.72

The results of aspect extraction are reported in Table III. The latter shows that the best performances are obtained in: Accuracy (84.26%), Recall (74%), Precision (77.14%) and F1-Score (75%) with the RandomForest + CountVectorizer configuration.

- Sentiment classification

TABLE IV. RESULTS OBTAINED FROM ALL CLASSIFIERS IN COUNTVECTORIZER WEIGHTING METHOD FOR SENTIMENT CLASSIFICATION

Metrics \ Classifiers	Accuracy %	Recall%	Precision %	F1-Score %
LR	<b>91</b>	<b>69.59</b>	<b>82</b>	<b>73.27</b>
RF	<b>88.87</b>	67.16	85	<b>72</b>
NB	83.28	60.43	67.26	61.58
DT	74.40	43.55	56.09	43.52
KNN	74.03	44.86	77.84	46.24
SVM	87.86	62.48	80.02	65.76
ET	88.53	68.19	82	72.59
AB	84.52	61.07	77.02	65.24
GB	87.03	57.64	58.58	57.85

The sentiment classification results are reported in Table IV. The latter shows that the best performances are obtained in precision (82%), recall (69.59%), accuracy (91%) and F1-score (73.27%) with the LogisticRegression + CountVectorizer configuration.

b) *TF-IDF Vectorizer embeddings classification*: Both Tables V and VI show the results of the classifier using the TF-IDF weighting model for sentiment classification.

- Aspect classification

TABLE V. RESULTS OBTAINED FROM ALL CLASSIFIERS IN TF-IDF WEIGHTING METHOD AOR ASPECT CLASSIFICATION

Metrics Classifiers	Accuracy %	Recall%	Precision %	F1-Score %
LR	83	70.01	78	72
RF	<b>86</b>	<b>74.78</b>	<b>81.01</b>	<b>76.59</b>
NB	76.59	72.08	50.55	51
DT	48.70	28.13	44.15	27
KNN	63.73	52.46	59.63	54.58
SVM	82.76	70.69	76.14	71.75
ET	82.80	72	76.04	73
AB	70	59.43	64.01	60.58
GB	71.63	52.04	52.42	51.44

The results of aspect classification with the TF-IDF vectorization method are presented in Table V. This latter shows that using the RandomForest + TF-IDF parameter provides the best performance in terms of precision (81.01%), recall (74.78%), accuracy (86%) and F1-score (76.59%).

- Sentiment classification

TABLE VI. RESULTS OBTAINED FROM ALL CLASSIFIERS IN TF-IDF WEIGHTING METHOD FOR SENTIMENT CLASSIFICATION

Metrics Classifiers	Accuracy %	Recall%	Precision %	F1-Score %
LR	87.37	60.73	83.84	64.04
RF	<b>89</b>	<b>66.19</b>	<b>84.51</b>	<b>70.55</b>
NB	82.87	53.17	57	54.02
DT	74.44	43.57	56.20	43.53
KNN	75.87	50.25	80	54.16
SVM	<b>87.63</b>	62.39	82.41	66.54
ET	<b>87.29</b>	76.26	78.37	71.10
AB	<b>84.82</b>	61.30	76.64	65.58
GB	<b>87.11</b>	57.50	59	57.89

The results of the sentiment classification are presented in Table VI. The latter shows that the use of the RandomForest + TF-IDF parameter allows to obtain the best performances in terms of accuracy (89%), Recall (66.19%), precision (84.51%) and F1-Score (70.55%).

## V. DISCUSSION

Traditional Sentiment Analysis is done through Sentiment Analysis techniques[36] on documents and sentences which assesses the overall polarity of the feelings of the given opinion target. Nevertheless, if the opinion target contains various aspects with a conflicting sentiment, using a single sentiment label to represent it could be incorrect[37]. The current trend is to move to a deeper level that presents itself as Aspect-Based Sentiment Analysis. ABSA is the sub-field of NLP that essentially breaks the data into aspects and finally extracts the sentiment information[38]. It performs a more advanced and higher quality analysis because it directly examines the sentiment itself. Neither document analysis nor sentence analysis find out what exactly people like and dislike. Specifically, the idea behind this work is to collect customer reviews on tourism sites such as: Booking or TripAdvisor, and assign a sentiment analysis that allows to extract the most relevant characteristics in the review of most customers. Hence the realization of a sentiment analysis as well as an aspect analysis on a dataset that contains customer reviews of hotels located in Marrakech. The results of this research were presented as follows:

The best result in all the tests for the classification of aspects is 86%, it was obtained by the RF with the use of TF-IDF, similarly the classifier GradientBoost, NB and AdaBoost reached their maximum measure (76.13%, 77%, 70% respectively), on the other hand for LR, SVM and KNN, their best results were with CountVectorizer (83.25%, 83.51%, 71.59%) respectively. Moreover, the DT classifier obtained the same result in both tests with TFIDF and CountVectorizer (48.70%). On the other hand, for sentiment classification the best result of Accuracy is 91% in CountVectorizer weighting method, achieved by LR classifier, also NB and SVM classifiers their excellent outcome were with the same method (83.28%, 87.86% respectively), also the DT classifier score is similar for both vectorization techniques with 74%, and both RF, GB, AdaBoost, and KNN classifiers got their excellent score with TF-IDF (89%, 87.11%, 84.82% and 75.87% respectively). Analyzing the results of the confusion matrices the RF+TFIDF that gave as results 2232 True Positives for the classification of the aspects, besides the best classifier for the analysis of the feelings is the LR+CountVectorizer that allows to reach a number of True Positives equal to 3158. These results show that RF is generally considered a better classifier for the aspect extraction task; in return LR is the good classifier for the sentiment classification task. As long as a good measurement is achieved, the model has permitted to perform correct results. These results are listed in Fig. 8 and 9.

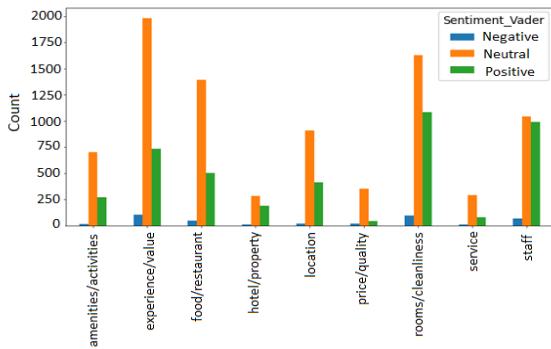


Fig. 8. Sentiment per single aspect

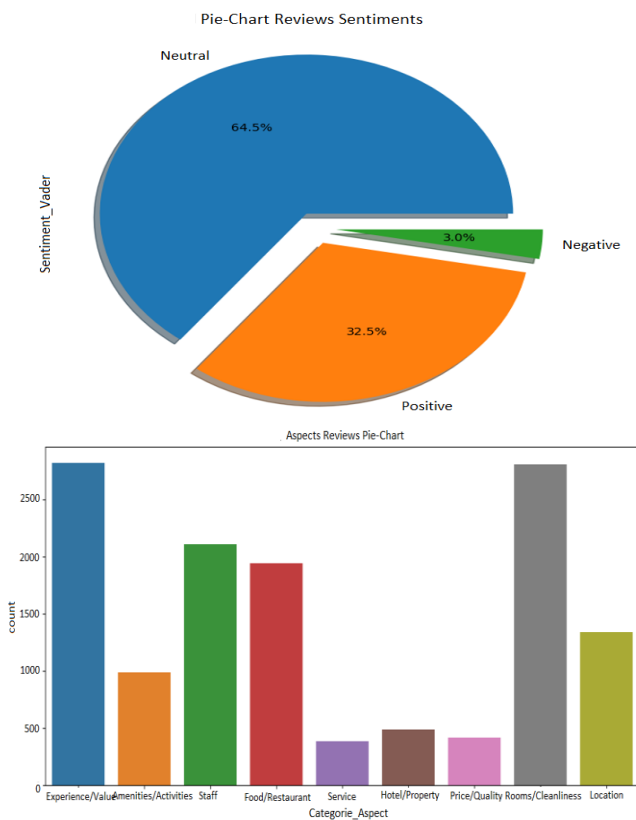


Fig. 9. Data analysis interface

## VI. CONCLUSION

Given the increasing importance placed on online reviews and the evidence that these reviews influence customer behaviors, it is clear that companies are beginning to look at technologies that could automatically analyze what customers are saying about their product or service. In this article, the analysis of the feelings has been performed on the level of aspects in the field of hotels. The hotel reviews that were studied are comments on the service provided, written in English on the rating site TripAdvisor. The preprocessing and vectorization methods are evaluated using nine supervised classification models. The results obtained are very encouraging, and the experimentation conducted on the dataset reveals that a better accuracy score of 92% and 3158 True Positives were achieved when using the LR+ CountVectorizer

classifier for sentiment analysis as well as a good Accuracy score of 86% and a number of True Positives equal to 2232 when using RF + TF-IDF Vectorizer for aspect classification. For future work, enriching the dataset with other reviews from other languages such as French and Arabic will be highly recommended and that is by making comparisons between these languages in order to get a broader view on the aspects that are most noticed by travelers from other countries; thus, the use of the mixed class analysis rather than the positive, negative and neutral classes.

## REFERENCES

- [1] C. Zong, R. Xia, and J. Zhang, Text Data Mining. Singapore: Springer Singapore, 2021. doi: 10.1007/978-981-16-0100-2.
- [2] A. D. Francisco, "Aspect Term Extraction in Aspect-Based Sentiment Analysis Aspect Term Extraction in Aspect-Based Sentiment Analysis," 2019.
- [3] I. E. Vermeulen and D. Seegers, "Tried and Tested: The Impact of Online Hotel Reviews on Consumer Consideration Tried and tested: The impact of online hotel reviews on consumer consideration," no. February 2009, 2020, doi: 10.1016/j.tourman.2008.04.008.
- [4] K. Ravi and V. Ravi, A survey on opinion mining and sentiment analysis: tasks, approaches and applications, no. June. Elsevier B.V., 2015. doi: 10.1016/j.knosys.2015.06.015.
- [5] J. Zelenka, T. Azubuike, and P. Martina, "administrative sciences Trust Model for Online Reviews of Tourism Services and Evaluation of Destinations," 2021.
- [6] H. A. Lee, R. Law, J. Murphy, and J. Murphy, "Helpful Reviewers in TripAdvisor, an Online Travel Community," no. September 2012, pp. 37–41, 2011, doi: 10.1080/10548408.2011.611739.
- [7] S. Shayaa, N. I. Jaafar, S. Bahri, A. Sulaiman, and M. A. L. I. Al-garadi, "Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges," IEEE Access, vol. 6, pp. 37807–37827, 2018, doi: 10.1109/ACCESS.2018.2851311.
- [8] V. Pekar and S. Ou, "Discovery of subjective evaluations of product," vol. 14, no. 2, pp. 145–155, 2008, doi: 10.1177/1356766707087522.
- [9] A. Muangon and S. Thammaboosadee, "A Lexiconizing Framework of Feature-based Opinion Mining in Tourism Industry," pp. 169–173, 2014.
- [10] E. Marrese-Taylor, J. D. Velasquez, and F. Bravo-Marquez, "OpinionZoom, a modular tool to explore tourism opinions on the Web," Proc. - 2013 IEEE/WIC/ACM Int. Jt. Conf. Web Intell. Intell. Agent Technol. - Work. WI-IATW 2013, vol. 3, pp. 261–264, 2013, doi: 10.1109/WI-IATW.2013.193.
- [11] Z. Hai, K. Chang, J. J. Kim, and C. C. Yang, "Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance," IEEE Trans. Knowl. Data Eng., vol. 26, no. 3, pp. 623–634, 2014, doi: 10.1109/TKDE.2013.26.
- [12] A. Mukherjee and B. Liu, "Aspect Extraction through Semi-Supervised Modeling," no. July, pp. 339–348, 2012.
- [13] W. Kasper and M. Vela, "Sentiment Analysis for Hotel Reviews," In Computational linguistics-applications conference, vol. 231527, pp. 45–52, 2011.
- [14] M. Pontiki et al., "SemEval-2016 Task 5: Aspect Based Sentiment Analysis," Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics, pp. 19–30, 2016.
- [15] A. Nazir, Y. Rao, L. Wu, and L. Sun, "Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey," no. October, 2021, doi: 10.1109/TAFFC.2020.2970399.
- [16] C. D. Manning, H. Schütze, and G. Weikurn, "Foundations of Statistical Natural Language Processing," SIGMOD Rec., vol. 31, no. 3, pp. 37–38, 2002, doi: 10.1145/601858.601867.
- [17] K. Mandal, "Topic Modeling: Techniques and AI Models - DZone," Dec. 15, 2020. <https://dzone.com/articles/topic-modelling-techniques-and-ai-models>.

- [18] B. Dutta, "What is Topic Modelling in NLP? Analytics Steps," Jan 15, 2022, <https://www.analyticssteps.com/blogs/what-topic-modelling-nlp>
- [19] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," vol. 3, pp. 993–1022, 2003.
- [20] T. K. Landauer and P. W. Foltz, "An Introduction to Latent Semantic Analysis," pp. 259–284, 1998.
- [21] Avinash Navlani, "Python LSI\_LSA (Latent Semantic Indexing\_Analysis) – DataCamp," Oct 2018. <https://www.datacamp.com/tutorial/discovering-hidden-topics-python>
- [22] D. D. Lee, M. Hill, and H. S. Seung, "Algorithms for Non-negative Matrix Factorization," *Adv. Neural Inf. Process. Syst.*, no. 1, pp. 556–562, 2001.
- [23] R. J. Gallagher, K. Reing, D. Kale, and G. Ver Steeg, "Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge," vol. 5, pp. 529–542, 2017.
- [24] M. Ounacer, S., Jihal, H., Ardchir, S., & Azzouazi, "Anomaly Detection in Credit Card Transactions," *Adv. Intell. Syst. Sustain. Dev.*, 2019, doi: 10.1007/978-3-030-36674-2\_14.
- [25] J. M. Hilbe, *Practical Guide to Logistic Regression* (1st ed.), Chapman and Hall/CRC, <https://doi.org/10.1201/b18678>, 2015.
- [26] V. Jakkula, "Tutorial on Support Vector Machine (SVM)," School of EECS, Washington State University. pp. 1–13, 2006.
- [27] Z. Zhang, "Introduction to machine learning : k-nearest neighbors," vol. 4, no. 11, pp. 1–7, 2009, doi: 10.21037/atm.2016.03.37.
- [28] I. Rish, "An empirical study of the naive Bayes classifier," *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, vol. 22, pp. 41–46, 2001.
- [29] A. Ashari, "Performance Comparison between Naïve Bayes , Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool," vol. 4, no. 11, pp. 33–39, 2013.
- [30] L. Breiman, "RANDOM FORESTS," *Mach. Learn.* 45, pp. 5–32, 2001, doi: <https://doi.org/10.1023/A:1010933404324>.
- [31] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," no. November 2005, pp. 3–42, 2006, doi: 10.1007/s10994-006-6226-1.
- [32] R. E. Schapire, P. Avenue, and A. Room, "The Boosting Approach to Machine Learning An Overview," pp. 1–23, 2003.
- [33] P. Celio, D. Cellio, S. Experian, M. Forti, M. Witorsa, and S. Experian, "A comparison of Gradient Boosting with Logistic Regression in Practical Cases GRADIENT BOOSTING MACHINES – THEORY," pp. 1–25, 2018.
- [34] V. Carle, "DEGREE PROJECT IN THE FIELD OF TECHNOLOGY Web Scraping using Machine Learning," 2020.
- [35] S. Ounacer, H. Jihal, K. Bayoude, A. Daif, and M. Azzouazi, "Handling Imbalanced Datasets in the Case of Credit Card Fraud," in *Advances in Intelligent Systems and Computing*, 2022, pp. 666–678. doi: 10.1007/978-3-030-90633-7\_56.
- [36] A. R. Alaei, S. Becken, and B. Stantic, "Sentiment Analysis in Tourism: Capitalizing on Big Data," *J. Travel Res.*, vol. 58, no. 2, pp. 175–191, 2019, doi: 10.1177/0047287517747753.
- [37] L. Zhu, M. Xu, Y. Bao, Y. Xu, and X. Kong, "Deep learning for aspect-based sentiment analysis: a review," *PeerJ Comput. Sci.*, vol. 8, 2022, doi: 10.7717/PEERJ-CS.1044.
- [38] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges," pp. 1–21, 2022, [Online]. Available: <http://arxiv.org/abs/2203.01054>



# Performance Comparison of the Kernels of Support Vector Machine Algorithm for Diabetes Mellitus Classification

Dimas Aryo Anggoro, Dian Permatasari

Informatics Engineering Department, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia

**Abstract**—Diabetes Mellitus is a disease where the body cannot use insulin properly, so this disease is one of the health problems in various countries. Diabetes Mellitus can be fatal and can cause other diseases and even lead to death. Based on this, it is important to have prediction activities to find out a disease. The SVM algorithm is used in classifying Diabetes Mellitus diseases. The purpose of this study was to compare the accuracy, precision, recall, and F1-Score values of the SVM algorithm with various kernels and data preprocessing. The data preprocessing used included data splitting, data normalization, and data oversampling. This research has the benefit of solving health problems based on the percentage of Diabetes Mellitus and can be used as material for accurate information. The results of this study are that the highest accuracy was obtained by 80% obtained from the polynomial kernel, the highest precision was obtained by 65% which was also obtained from the polynomial kernel, and the highest recall was obtained by 79% obtained from the RBF kernel and the highest f1-score was obtained by 70% obtained from RBF kernel.

**Keywords**—Diabetes mellitus; kernel; normalization; oversampling; SVM

## I. INTRODUCTION

Diabetes Mellitus is a disease where blood sugar levels are overly high because the body cannot use insulin properly. Currently, Diabetes Mellitus becomes a serious health problem in various countries, including Indonesia [1]. The International Diabetes Federation (IDF) explained that in 2021 the number of people with Diabetes Mellitus in Indonesia reached 19.5 million people, while in 2019 the figure was 10.7 million. This means that there has been an increase of nearly 9 million cases in just 2 years, or just during the COVID-19 pandemic. With almost 2 times the addition, makes Indonesia ranked fifth in the world. Not only in Indonesia, but this upward trend in cases also occurs in the world. According to IDF data, currently, at least 1 in 10 people or as many as 537 million people in the world live with Diabetes Mellitus. If not treated properly immediately, Diabetes Mellitus can be fatal and can cause other diseases and even lead to death. Based on this, it is important to have prediction activities to find out a disease. This activity is carried out so that a disease can be detected quickly and can be treated immediately.

Activities in predicting various diseases have been carried out in various scientific fields, one of which is the field of computer science. Along with the development of information and communication technology, it can be used to improve the

ability of the system to help detect Diabetes Mellitus disease[2]. Data mining is part of the Knowledge Discovery in Database (KDD) process that can classify, predict, and get a lot of information from large data sets[3]. Classification is an important stage in data mining; classification is carried out by looking at variables from existing data groups and aims to predict the class of an object that was not previously known [4].

## II. LITERATURE REVIEW

Previous research conducted by Andi Maulida Argina regarding the application of the K-Nearest Neighbour classification model to the diabetes patient dataset explained that the study had the highest accuracy of 39%[5]. Another study conducted by Noviandi on the implementation of the Decision Tree C4.5 algorithm for diabetes prediction resulted in a prediction model that had the highest accuracy of 70.32% [6]. The shortcoming of the previous study is that the accuracy of the prediction model is still below 80%, so there is a need to improve accuracy performance. In research [7] that compared accuracy, recall, and precision classification on the C4.5 algorithm, Random Forest, Support Vector Machine (SVM), and Naïve Bayes resulted in the C4.5 algorithm obtaining accuracy of 86.67%, the Random Forest algorithm obtained accuracy of 83.33%, the SVM algorithm obtained accuracy by 95%, and the Naive Bayes algorithm obtained an accuracy of 86.67%. The highest accuracy algorithm is the SVM algorithm, therefore in this study applying the SVM algorithm for the classification of Diabetes Mellitus disease. This research is expected to provide accuracy results reaching 80%, so that it can improve deficiencies in previous studies.

The SVM algorithm was chosen because it is reliable in processing large amounts of data by optimizing hyperplanes in high-dimensional space that maximizes margins between data [8]. The use of the kernel in SVM is carried out to determine kernel parameters and produce the best accuracy in the classification process. Linear kernels are used when classified data can be easily separated by a hyperplane, while non-linear kernels are used when the data used is separated using curved lines or a plane in space that has high dimensions [9].

This study aims to compare the accuracy, precision, recall, and F1-Score values of the SVM algorithm with various kernels and preprocessing data in the classification of Diabetes Mellitus disease. It has the benefit of solving health problems based on the percentage of Diabetes Mellitus and can be an accurate information material. The output of this study is to

imply that the SVM algorithm is expected to show better performance values than previous studies.

### III. METHODOLOGY

#### A. Data Collection

The first stage in this study is the collection of Diabetes Mellitus datasets. The dataset used is the Pima Indian diabetes dataset obtained from the UCI Machine Learning Repository. Several variables and attributes can facilitate the research process in data mining. The Pima Indian diabetes dataset consists of 768 data and 9 attributes. The variables and attributes used are shown in Table I.

TABLE I. VARIABLES AND ATTRIBUTES OF PEOPLE WITH DIABETES MELLITUS

Variable	Attribute
X1	<i>Pregnancies</i> , the number of pregnancies during life in the range of 0-17 times.
X2	<i>Glucose</i> , glucose/blood sugar levels. Normal blood sugar levels are below 120 mg/dL, while the sugar levels of diabetics are more than 120 mg/dL. The data range in the dataset is 0-199 mg/dL.
X3	<i>Blood Pressure</i> , blood pressure with mmHg units, the data range in the dataset is 0-112 mmHg.
X4	<i>Skin Thickness</i> , skin fold thickness with a data range of 0-99 mm. The norm is about 12.5 mm.
X5	<i>Insulin</i> , insulin levels in the blood with a data range of 0-846 U / ml.
X6	<i>BMI</i> , body mass weight with a data range of 0-67.1 BMI
X7	<i>Diabetes Pedigree Function</i> , History of diabetes Mellitus disease in the family with a data range of 1.001-2.42.
X8	<i>Age</i> , age of the patient (years) with a data range of 21-81 years.
Y	<i>Outcome</i> , negative and positive class variables (0 and 1). 0 are indicators of non-diabetics while 1 is an indicator of diagnosed diabetics.

#### B. Data Preprocessing

1) *Data splitting*: The next stage is the data splitting stage which is carried out by separating training data and testing data. Training data is used to create models that are applied to testing data [10] and testing data cannot be used for the training process, so that the model really learns from the new data [11]. The determination of training data and testing data is carried out randomly, so that the proportion between categories remains balanced [12]. In this study, splitting data was divided into 80% training data and 20% testing data.

2) *Normalization data*: Normalization of data in datasets aims to create data in the same range of values [13]. This study used the min-max and z-score normalization methods.

a) *Min-max normalization*: Normalization of min-max can overcome non-uniform data forms with a range of values greater than 0-1 [14]. Min-max normalization was chosen because it has the advantage that the data is balanced between before and after normalization [15]. The normalization of min-max is presented in (1).

$$x_{new} = \frac{x_{old} - x_{min}}{x_{max} - x_{min}} \quad (1)$$

$x_{new}$  represents the min-max value,  $x_{old}$  is the value to be normalized,  $x_{min}$  is the lowest value of the overall data and  $x_{max}$  is the highest value of the entire data.

b) *Z-Score normalization*: Z-Score normalization is used to compare the performance or quality of data goals with the average distribution of data across groups based on standard deviation values [16]. Z-score normalization was chosen because it is a good normalization method for balancing data scale [17]. (2) is a formula for knowing the z-score.

$$x_{new} = \frac{x_{old} - \mu}{\sigma} \quad (2)$$

$x_{new}$  is the z-score value,  $x_{old}$  is the value to be normalized,  $\mu$  is the average value of the whole data and  $\sigma$  is the standard deviation value.

#### 3) Oversampling

a) *SMOTE (Synthetic Minority Over-sampling Technique)*: The SMOTE method can handle dataset class imbalances by working to make data replication of minor classes to be equivalent to major classes [18]. The diabetes dataset used in this study had a total of 268 positive classes and 500 negative classes so that there was an imbalance between the positive class and the negative class. Therefore, the SMOTE method was used in this study to balance between positive classes and negative classes. (3) is the formula for SMOTE.

$$x_{syn} = x_i + (x_{knn} - x_i)\gamma \quad (3)$$

$x_{syn}$  is the resulting new class data,  $x_i$  is the approach to i,  $x_{knn}$  is the x closest to  $x_i$  and  $\gamma$  is a random number between 0-1.

#### C. Data Processing

1) *Support Vector Machine (SVM)*: SVM is a good algorithm for data classification [19] with the principle of finding the best hyperplane that serves as a separator of two data classes [20]. The best hyperplane is determined by measuring the hyperplane margin and finding its maximum point, margin is the distance between the hyperplane and the nearest point of each class and this closest point is called the support vector [21]. The following is a description of SVM, there is data  $\vec{x}_i \in (\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n)$   $x_i$  is data consisting of n attributes and two classes  $y_i \in +1, -1$ . Suppose that the two classes can be perfectly separated by a d-dimensional hyperplane defined by (4).

$$\vec{w} \cdot \vec{x}_i + b = 0 \quad (4)$$

Data  $\vec{x}_i$  which belonging to the positive class (+1) are shown in (5).

$$\vec{w} \cdot \vec{x}_i + b \geq -1 \quad (5)$$

Meanwhile, data  $\vec{x}_i$  belonging to the negative class (-1) are shown in (6).

$$\vec{w} \cdot \vec{x}_i + b \leq +1 \quad (6)$$

The maximum margin can be obtained by maximizing the value of the distance between the hyperplane and its closest point or support vector which  $\frac{1}{\|\bar{w}\|}$  [22]. It is formulated as Quadratic Programming (QP) by looking for a minimum point based on (7).

$$\min \tau(w) = \frac{1}{2} \|\bar{w}\|^2 \quad (7)$$

By paying attention to the constraints on (8).

$$y_i(\bar{x}_i, \bar{w} + b) - 1 \geq 0 \quad (8)$$

$y_i$  is the target class to  $i$ ,  $\bar{x}_i$  is the input data to  $i$ ,  $\bar{w}$  is the weight, and  $b$  is the relative field position.

2) *Kernel SVM*: To work around high-dimensional data can use a kernel that transforms the input space into a feature space [23]. Kernel functions commonly used in SVM are Linear [24], Radial Basic Function (RBF) and Polynomial [25]. The parameters possessed by kernel functions are used in the testing process [26]. There is no definite conclusion about the best kernel, therefore this study will compare 4 kernel functions, namely linear, RBF, polynomial and sigmoid.

a) *Kernel linear*: The Linear kernel was chosen because it is the simplest kernel and is used when the data is linearly overstretched.

$$K(x, y) = x \cdot y \quad (9)$$

b) *Kernel polynomial*: The Polynomial kernel was chosen because it can be used when the data is not linearly separated and is suitable for use in solving classification problems in all training data that has been normalized.

$$K(x, y) = (\gamma(x \cdot y) + C)^d \quad (10)$$

3) *Kernel Radial Basic Function (RBF)*: The RBF kernel is used when the data is not linearly separated, it is chosen because it performs well with certain parameters, and the result of the training has a small error value.

$$K(x, y) = \exp(-\gamma|x - y|^2) \quad (11)$$

a) *Kernel sigmoid*: This sigmoid kernel was chosen because it is similar to the two-layer perceptron model of the neural network, which works as an activation function for neurons.

$$K(x, y) = \tanh(\gamma(x \cdot y) + C) \quad (12)$$

4) *Evaluation*: Confusion matrix is an evaluation method that provides information comparing the classification of prediction results with the actual classification [27]. There are 4 terms of value from the confusion matrix, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Based on these values, accuracy, precision, recall, and F1-Score values can be generated. Accuracy is the ratio of predicted correct values of all data [28].

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (13)$$

Precision indicates a correctly classified prediction of positive values divided across positive classified data [28].

$$Precision = \frac{TP}{TP+FP} \quad (14)$$

Recall shows the comparison of the positive correct predicted value with the entire positive correct value [29].

$$Recall = \frac{TP}{TP+FN} \quad (15)$$

The F1-Score shows the average comparison of precision and recall values [29].

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (16)$$

## IV. RESULTS AND DISCUSSIONS

This stage is a decipherment of the research obtained and its explanation.

### A. Data Preprocessing

The dataset used is the Pima Indian diabetes dataset which consists of 768 data and 9 attributes. The initial stage carried out in this study is the process of collecting and processing datasets. In this study, data preprocessing was divided into three steps. The first step is the data splitting process, where the Diabetes Mellitus dataset will be divided into training data and testing data. The second step is the data normalization process to create data in the same range. The third step is an oversampling process to balance the dataset class by using the SMOTE method. Data processing in the study uses the Python programming language in the Google colab application.

1) *Data splitting results*: After getting the dataset, the next step is to divide the dataset into training data and testing data. The Diabetes Mellitus dataset totaled 768 data consisting of 8 variables and one target/class. Then the dataset is divided into 80% training data, totaling 614 data and 20% testing data, totaling 154 data. The diagnosis of Diabetes Mellitus is divided into two, namely non-diabetics who are denoted by 0 and diabetics who are denoted by 1. Obtained diabetics totaled 268 data and non-diabetics amounted to 500 data.

2) *Data normalization results*: The normalization methods used are min-max and z-score.

Fig. 1 shows the comparison of variables in the dataset, variables compared to 2, namely, pregnancies and insulin, the data has a fairly high range of values. For example, in the insulin variable, where the range of values is between 0 to above 200, this is considered unbalanced. The min-max normalization method is used to process values into the range of 0-1. Fig. 2 shows the results after normalization of min-max, where the range of values in the insulin variable becomes smaller, namely, 0-1.

In addition to using the min-max method, data normalization is also carried out using the z-score method. Z-score is performed by processing the mean and standard deviation from the values of its attributes. Fig. 3 shows the results after normalizing the z-score.

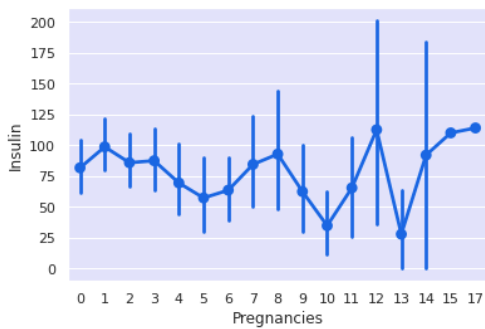


Fig. 1. Before normalization.



Fig. 2. After min-max normalization.



Fig. 3. After z-score normalization.

3) *Oversampling results:* In the dataset there is a difference between the number of positive classes and negative classes, therefore there is a need for class balancing. Class balancing is done by oversampling using the SMOTE method and is carried out on training data only. Oversampling is carried out after splitting data so that data replication does not appear in data training and data testing [30]. It can be seen in Fig. 4, before oversampling the number of positive classes was 221 and the number of negative classes was 393. Meanwhile, after oversampling, the number between the positive class and the negative class becomes the same, which is 393 so that it becomes balanced.

### B. Data Preprocessing and Evaluation

This study compared the performance of the SVM algorithm kernels for the classification of Diabetes Mellitus diseases. SVM kernels used include linear kernels, polynomial kernels, RBF kernels, and sigmoid kernels. Evaluation is carried out using the confusion matrix method to calculate the values of accuracy, precision, recall, and f1-score by optimizing the best parameters for each kernel. Each kernel on SVM has a specific parameter, the cost parameter (C) being the most commonly used value for all kernels. The gamma ( $\gamma$ )

parameter is used to determine the degree of proximity between 2 points to make it easier to find  $\gamma$  hyperplanes that are consistent with the data. The gamma parameter is used by polynomial, RBF, and sigmoid kernels. Next is the degree (d) parameter used to map data from the input space to the higher dimension space in the feature space, only the polynomial kernel uses this parameter [31]. Determination of the best parameters on the kernel is carried out by trial and error. Table II is the result of evaluating the classification models of various SVM kernels before various data preprocessing is carried out.

For this experiment in Table II, all parameter values in each kernel use auto parameters from python. The highest accuracy is obtained from the polynomial and RBF kernels, which is 77%. The highest precision was obtained from the RBF kernel, which was 69%, the highest recall was obtained from the linear kernel, which was 57% and the highest f1-score was obtained from linear and polynomial kernels, which was 61%. Table III is the result of evaluating the classification models of various SVM kernels after preprocessing data with min-max normalization and SMOTE oversampling. Meanwhile, Table IV is the result of evaluating the classification models of various SVM kernels after preprocessing data with normalization of z-score and oversampling SMOTE.

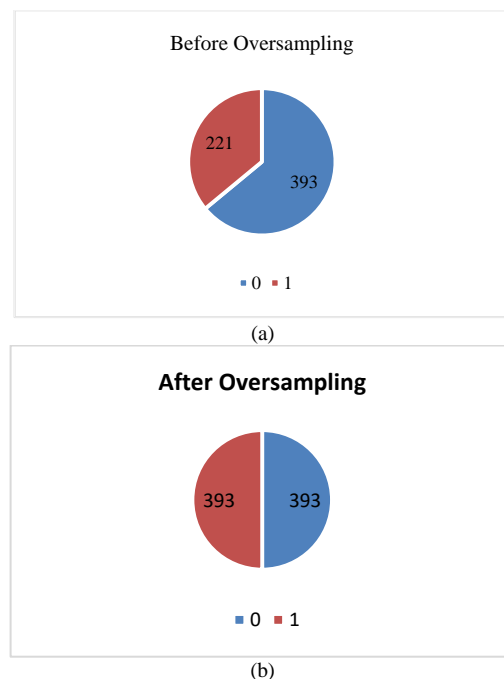


Fig. 4. (a) Data before oversampling, (b) Data after oversampling.

TABLE II. RESULTS OF EVALUATION OF VARIOUS SVM KERNELS BEFORE DATA PREPROCESSING

	Kernel			
	Linear	Polynomial	RBF	Sigmoid
Accuracy	76%	77%	77%	51%
Precision	66%	68%	69%	12%
Recall	57%	55%	53%	8%
F1-Score	61%	61%	60%	10%

TABLE III. RESULTS OF EVALUATION WITH MIN-MAX AND SMOTE

	Kernel			
	Linear	Polynomial	RBF	Sigmoid
Accuracy	77%	79%	79%	76%
Precision	61%	64%	63%	59%
Recall	72%	72%	79%	68%
F1-Score	66%	68%	70%	63%

TABLE IV. RESULTS OF VALUATION WITH Z-SCORE AND SMOTE

	Kernel			
	Linear	Polynomial	RBF	Sigmoid
Accuracy	79%	80%	77%	78%
Precision	62%	65%	61%	62%
Recall	74%	74%	72%	72%
F1-Score	68%	69%	66%	67%

Based on Tables III and IV, it can be seen that the highest accuracy is obtained by applying z-score normalization and SMOTE oversampling, which is obtained by 80% using a polynomial kernel. The polynomial kernel using the parameter value  $C=1$   $\gamma=0.1$   $d=1.5$  is obtained through trial and error so that it can produce margin optimization values that are used to maximize the hyperplane by mapping the data into higher dimensions. The highest precision is also obtained from the polynomial kernel, which is 65%. This shows that the higher the accuracy value, the higher the precision value will be. The highest recall was obtained at 79% which is from the RBF kernel shown in Table III. The RBF kernel uses the parameter value  $C=2.5$   $\gamma=1.5$ . The highest F1-score is also obtained from the RBF kernel shown in Table III which is 70%. The values in the parameters C,  $\gamma$ , and d are the most optimal values in order to get the maximum accuracy value. If the value is increased or decreased, the accuracy value will decrease.

## V. CONCLUSION AND FUTURE WORK

This research produces the highest accuracy of up to 80% which is obtained from polynomial kernels. So that the shortcomings of previous research have been resolved in this study. By optimizing the use of the kernel on the SVM algorithm it is proven to be able to maximize performance. So it can be concluded that the SVM algorithm shows a better performance value in classifying Diabetes Mellitus. Where in this study it was found that the performance of the SVM algorithm kernel to produce the highest accuracy was obtained from the polynomial kernel. The accuracy produced in this study can be used as an accurate and useful information material for overcoming health problems based on the percentage of Diabetes Mellitus.

For further research, you can use other datasets that have more data and also use other algorithms such as Xgboost, Bayesian Classification and other algorithms to get better accuracy. In addition, the results of this study can also be used in making applications to detect Diabetes Mellitus which can be web-based or mobile.

## REFERENCES

- [1] N. Shamsiyah, "Mengenal diabetes melitus." In *Berdamai dengan diabetes*, pp. 1-12. Jakarta : Bumi Media, 2022.
- [2] S. Wiyono, "Perbandingan kinerja rule zeroR dan function SMO dengan T-test dalam pengklasifikasian diagnosis penyakit diabetes mellitus." *Jurnal Teknik Elektro*, vol. 16, no. 01, pp. 23–25, 2016.
- [3] Y. Mardi, "Data mining : Klasifikasi menggunakan algoritma C4.5." *Jurnal Edik Informatika*, vol. 2, no. 2, pp. 213–219, 2017.
- [4] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan normalisasi data untuk klasifikasi wine menggunakan algoritma K-NN." *CESS (Journal of Computer Engineering System and Science)*, vol. 4, no. 1, pp. 78–82, 2019.
- [5] A. M. Argina, "Penerapan metode klasifikasi k-nearest neighbor pada dataset penderita penyakit diabetes." *Indonesian Journal of Data Sciencs*, vol. 1, no. 2, pp. 29–33, 2020.
- [6] Noviandi, "Implementasi algoritma decision tree C4.5 untuk prediksi penyakit diabetes." *Jurnal INOHIM*, vol. 6, no. 1, pp. 1–5, 2018.
- [7] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan akurasi, recall, dan presisi klasifikasi pada algoritma C4.5, random forest, SVM dan naive bayes." *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, pp. 640–651, 2021.
- [8] Y. Zhao, C. Zhang, Y. Zhang, Z. Wang, and J. Li, "A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis." *Energy and Built Environment*, vol. 1, no. 2, pp. 149–164, 2020.
- [9] U. P. Harapan, D. E. Ratnawati, and A. W. Widodo, "Klasifikasi penyakit gigi dan mulut menggunakan metode support vector machine." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 2 2018.
- [10] A. F. Rina Kurniasari, "Penerapan algoritma C4.5 untuk penjurusan siswa sekolah menengah atas." *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, vol. 8, no. 1, 2019.
- [11] B. A. H. and F. A. S. B. Helmi Imaduddin, "Arison of support vector machine and decision tree methods in the classification of breast cancer." *Jurnal Pendidikan Teknologi Informasi*, vol. 5, pp. 22–30, 2021.
- [12] R. A. Helena Nurramdhani Irmanda, "Klasifikasi jenis pantun dengan metode support vector machines (SVM)." *JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 5, pp. 915–922, 2020.
- [13] D. M. Ahmad Harmain, Paiman, Henri Kurniawan, Kusri, "Normalisasi data untuk efisiensi k-means pada pengelompokan wilayah berpotensi kebakaran hutan dan lahan berdasarkan sebaran titik panas." *TEKNIMEDIA*, vol. 2, no. 2, pp. 83–89, 2021.
- [14] H. E. Wahanani, M. H. P. Swari, and F. A. Akbar, "Case based reasoning prediksi waktu studi mahasiswa menggunakan metode euclidean distance dan normalisasi min-max." *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 6, pp. 1279–1288, 2020.
- [15] R. Fatwa, I. Cholissodin, and Y. A. Sari, "Penerapan metode extreme learning machine untuk prediksi konsumsi batubara sektor pembangkit listrik tenaga uap." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 11, pp. 10749–10755, 2019.
- [16] U. Al, A. Mandar, and S. Basri, "Novelty ranking approach with z-score and fuzzy multi-attribute decision making combination." *International Journal of Engineering & Technology*, vol. 7, no. 7, pp. 476–480, 2018.
- [17] T. M. Fahrudin, P. A. Riyantoko, K. M. Hindrayani, and M. H. P. Swari, "Cluster analysis of hospital inpatient service efficiency based on BOR, BTO, TOI, AvLOS indicators using agglomerative hierarchical clustering." *Telematika*, vol. 18, no. 2, p. 194, 2021.
- [18] R. Siringoringo, "Klasifikasi data tidak seimbang menggunakan algoritma SMOTE dan k-nearest neighbor." *Journal Information System Development (ISD)*, vol. 3 no. 1, 2018.
- [19] N. Nurajijah, D. A. Ningtyas, and M. Wahyudi, "Klasifikasi siswa SMK berpotensi putus sekolah menggunakan algoritma decision tree, support vector machine dan naive bayes." *Jurnal Khatulistiwa Informatika*, vol. 7, no. 2, 2019.
- [20] H. Nalatissifa, W. Gata, S. Diantika, and K. Nisa, "Perbandingan kinerja algoritma klasifikasi naive bayes, support vector machine (SVM), dan

- random forest untuk prediksi ketidakhadiran di tempat kerja." *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 4, p. 578, 2021.
- [21] A. A. Kasim and M. Sudarsono, "Algoritma support vector machine (SVM) untuk klasifikasi ekonomi penduduk penerima bantuan pemerintah di kecamatan simpang raya sulawesi tengah." *SEMNASITIK*, pp. 568–573, 2019.
- [22] S. A. Naufal, A. Adiwijaya, and W. Astuti, "Analisis perbandingan klasifikasi support vector machine (SVM) dan k-nearest neighbors (KNN) untuk deteksi kanker dengan data microarray." *JURIKOM (Jurnal Riset Komputer)*, vol. 7, no. 1, pp. 162–168, 2020.
- [23] S. Widodo, R. N. Rohmah, B. Handaga, L. Dyah, and D. Arini, "Lung diseases detection caused by smoking using support vector machine." *TELEKOMUNIKA*, vol. 17, no. 3, pp. 1256–1266, 2019.
- [24] R. Wati and S. Ernawati, " Analisis sentimen persepsi publik mengenai PPKM pada twitter berbasis SVM menggunakan python." *Jurnal Pendidikan Teknologi Informasi*, vol. 06, pp. 240–247, 2021.
- [25] E. Anindika Sari, M. Thereza Br. Saragih, I. Ali Shariati, S. Sofyan, R. Al Baihaqi, and R. Nooraeni, "Klasifikasi kabupaten tertinggal di kawasan timur indonesia dengan support vector machine." *JIKO (Jurnal Informatika dan Komputer)*, vol. 3, no. 3, pp. 188–195, 2020.
- [26] R. H. Muhammadiyah, T. G. Laksana, and A. B. Arifa, "Combination of support vector machine and lexicon-based algorithm in twitter sentiment analysis." *Khazanah Informatika: Journal Ilmu Komputer dan Informatika*, vol. 8 no. 1, 2021.
- [27] M. Syukron, R. Santoso, and T. Widiarihi, "Perbandingan metode smote random forest dan smote xgboost untuk klasifikasi tingkat penyakit hepatitis C pada imbalance class data." *Jurnal Gaussian*, vol. 9, no. 3, pp. 227–236, 2020.
- [28] M. Rangga, A. Nasution, and M. Hayaty, " Perbandingan akurasi dan waktu proses algoritma K-NN dan SVM dalam analisis sentimen twitter." *Jurnal Informatika*, vol. 6, no. 2, pp. 212–218, 2019.
- [29] A. Ridhovan, A. Suharso, "Penerapan metode residual network (RESNET) dalam klasifikasi penyakit pada daun gandum." *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 7, no. 1, pp. 58–65, 2022.
- [30] A. A. Arifiyanti and E. D. Wahyuni, "Smote : Metode penyeimbangan kelas pada klasifikasi data mining." *SCAN-Jurnal Teknologi Informasi dan Komunikasi*, vol. 15 no. 1, pp. 34–39, 2020.
- [31] I. M. Yulietha and S. Al Faraby, "Klasifikasi sentimen review film menggunakan algoritma support vector machine." *eProceedings of Engineering*, vol. 4, no. 3, pp. 4740–4750, 2017.

# Image Segmentation of Intestinal Polyps using Attention Mechanism based on Convolutional Neural Network

Xinyi Zheng<sup>1</sup>, Wanru Gong<sup>2</sup>, Ruijia Yang<sup>3</sup>, Guoyu Zuo<sup>4\*</sup>

Faculty of Information Technology, Beijing University of Technology, Beijing 100083, China<sup>1,2,4</sup>  
School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China<sup>3</sup>

**Abstract**—The intestinal polyp is one of the common intestinal diseases, which is characterized by protruding lining tissue of the colon or rectum. Considering that they may become cancerous, they should be removed by surgery as soon as possible. In the past, it took a lot of manpower and time to identify and diagnose intestinal polyps, which greatly affected the treatment efficiency of medical staff. Because the polyp part looks similar to the normal structure of the human body, the probability of human eye misjudgment is high. Therefore, it is necessary to use advanced computer technology to segment the intestinal polyp image. In the model established in this paper, an image segmentation method based on convolution neural network is proposed. The Har-DNet backbone network is used as the encoder in the model, and its feature processing results are converted into three feature images of different sizes, which are input to the decoding module. In the decoding process, each output first expands the receptive field module and then fuses the feature image processed by the attention mechanism. The fusion results are input to the density aggregation module for processing to improve the operation efficiency and accuracy of the model. The experimental results show that compared with the previous Pra-Net model and Har-DNet MSEG model, the accuracy and precision of this method are greatly improved, and can be applied to the actual medical image recognition process, thus improving the treatment efficiency of patients.

**Keywords**—Image segmentation; intestinal polyps; block convolutional attention mechanism; Har DNet

## I INTRODUCTION

Rectal polyps generally refer to the tissues protruding from the surface of the rectal mucosa to the intestinal cavity, which is also a high-risk factor for rectal cancer. Studies have pointed out that 95% of rectal cancer is caused by colorectal adenocarcinoma. To effectively avoid the incidence of rectal cancer, regular colonoscopy or resection of rectal polyps are effective means. During these years, with medical knowledge popularization and publicity about the causes of rectal cancer and rectal polyps, more and more people are aware of the importance of regular colonoscopy. This medical examination also depends on the doctor's own experience and competence, so it is a very arduous task for mankind. As an item of physical examination, a manual colonoscopy is a great test for doctors' clinical experience and ability. Most doctors need a lot of time to identify intestinal polyps, and also need a lot of experience to ensure the correctness of the results. Therefore, manual examination of intestinal polyp images by medical personnel is

very time-consuming work and depends on some experienced doctors. Therefore, people begin to use computers to process images instead of human to identify and segment intestinal polyps by combining image processing technology in the computer field to develop its application in the medical field.

In the medical science field, image segmentation technology is the basis of image processing in the relevant area of computers. Medical imaging (radiology) is a field where medical staff recreates images for diagnosis and treatment purposes. In order to improve the efficiency and accuracy of judging the pathological position, some valuable methods for information processing in the computer science area are adopted. After the computer processes the image and then crops the images with the elements such as brightness and texture in the image as the standard for segmenting, the model of image segmentation through a neural network has been more clearly established. Moreover, as the model is constantly improved with the algorithm, together with the adopting of a convolution neural network to process the image, the quality of the outcome has been greatly improved.

This model also applies the method of image segmentation in computer science. The core is the joint action of the convolution Neural Network (CNN) and Convolutional Block Attention Module (CBAM). The selected intestinal polyp data set is the actual medical image of the human intestinal cavity, and the polyp part has similar surface characteristics to the normal structure of the human body, which is not easy to distinguish. Therefore, segmentation of intestinal polyp images demands both the accuracy and precision of computer image processing. This model chooses to enhance the feature extraction ability to strengthen the feature of the intestinal polyp feature different from the normal intestinal cavity tissue of the human body. This model reduces the possibility of misdiagnosis in the process of computer processing by enhancing the feature extraction ability, and reduces the time of feature processing by feature enhancement, so as to jointly improve the performance of the model in terms of time cost and accuracy. The choice of CBAM mainly depends on the difficulty of intestinal polyp image processing. Through the joint processing of the attention mechanism and receptive field expansion module, the intestinal polyp features of the image are effectively enhanced, which provides excellent pre-processing for feature extraction and image segmentation.

\*Corresponding Author.

In the part of image feature extraction, this model selects the Har DNet convolution network with perfect function as the backbone network in the encoder stage. However, due to the full integration of upper and lower image features by Har DNet, the amount of calculation required to output the feature image is large. Although the accuracy and effectiveness of segmented image features are retained to a great extent, there is still room for improvement in processing speed. This model improves the processing speed based on Har DNet. Through the expansion of the receptive field and the addition of an attention mechanism, it not only improves the accuracy of feature processing but also accelerates the processing, so as to effectively improve the performance of overall image processing.

A more rapid and precise processing of the images of intestine polyps has a huge and profound impact on the medical world. In an age when computers were not used to help with medical images processing, homeopathy had to invest in a number of people in the determination of the images of intestine polyps. This does not only require the abilities of doctors, but also the time and effort of the medical personnel. This paper proposes and uses an image-processing model that can free therapeutic staff in a maximum degree. With the aid of computer technology, medical personnel do not have to spend much time identifying the pathological tissue that is very close to normal tissue. The doctor can devote more energy to the treatment of the follow up. In addition, considering the high accuracy of the model, when small colonic polyps are encountered, the model can still accurately distinguish to achieve the effect of early warning.

This paper will start with the evolution of various neural network models, emphasize the performance improvement of numerous models over time, and lead to the optimization of subsequent models in this paper. After that, the main structure of the model is the core content of the article, which is described in combination with the structure diagram. Finally, this paper will list the results of relevant experiments for readers to intuitively experience the performance improvement of this model.

## II RELATED WORK

CNN has achieved great success in the field of medical image segmentation because of its excellent feature extraction ability and good feature expression ability [1,7], and it does not need manual extraction or much pre-processing work. The U-Net neural network [2-5] is one of the earliest adopted models in the semantic segmentation network [26]. Due to the data augmentation by elastic distortions, it only needs very few annotated images. However, since each pixel needs to take an image block [17] centered on itself, two adjacent pixels are highly similar in their block information, which causes much redundancy and slow network training. Therefore, on this basis, the U-Net++ network [14-15] is developed. Through the effective integration of U-Net in different depths, that is, these U-Net parts share an encoder, and effectively recover the fine-grained details of the target object under complex background utilizing in-depth supervision and joint learning. In short, U-Net++ is equivalent to splicing four U-Net networks with different depths through multiple skip connections. The first

advantage is the improvement of accuracy, which should be brought by integrating the characteristics of different levels. The second is the flexible network structure combined with deep supervision so that the deep network with huge parameters can greatly reduce the parameters within the acceptable accuracy range, which has been developed to more advanced models [27].

The ResUNet model [18-19] performs better in-depth and avoids the degradation of the network through residual learning. The main idea of this model is to add a direct channel in the network and control the neural network of the receiving layer by allowing the original input information to be directly transmitted to the later layer. Instead of learning all the outputs after complete calculation, it can learn the residual of the output of the previous network [19]. This kind of information processing protects the integrity of information. The whole network only needs to learn the part of the difference between input and output, simplify the learning objectives and difficulties, and significantly improve the output results.

The PraNet model [9], which draws reverse attention, can effectively improve the style progress and accurate characterization of the results by mining boundary clues through the reverse attention module and establishing the connections between regions and boundaries. This model also gives inspiration for our model: the establishment of the attention module helps to build the relationship between the actual image pixels. The attention mechanism can be used as a weighting module to basically process the features in the image before processing the image so that the features in the image have a judgment similar to "important" or "unimportant". This judgment has high accuracy because of the strong correlation of the information around the attention mechanism's domain. As a kind of attention mechanism, the block attention mechanism proposed in this model can help the model give different weights to each part of the input and extract the more critical information so that the model can make a more accurate judgment. It is a lightweight attention mechanism module, so it will not bring more overhead to the calculation and storage of the model and improves the accuracy and calculation speed of the model from multiple layers.

HarDNet, as an improved model of L2-Net, measures the amount of data accessing memory through convolution input-output (CIO) [10-11]. The amount of data in storage that can be accessed through HarDNet can improve the density of calculation. Compared with the previous convolutional neural network [6,8], it has a very significant improvement in the accuracy and speed of calculation.

In recent years, the HSNet [28] model has emerged as a new neural network model for image segmentation. The model uses a convolutional neural network (CNN) to predict the 3D model of the human body, thus avoiding the limitations of handmade features and postures. At the same time, it is applied in virtual fitting and human body size measurement. In the paper on the HSNet model, four possible situations are analyzed as the input of the network: (a) The single binary contour of a human is scaled to a fixed size to prevent the loss of camera calibration information; (b) The human shadow image is scaled to a fixed size because the shadow will retain



the information complementary to the contour; (c) Assume the front profile of known camera parameters; (d) Assume that the front and side profiles of the camera parameters are known. HSNNet [28] uses CNN to accurately predict 3D human models from contour or shadow images and tries to find the global mapping of deformation parameters from 2D images to 3D models. The model has been widely evaluated and tested on thousands of people and real people. In addition, this model has been proved by comprehensive experiments that better prediction results can be obtained if there is shadow information. The model further assumes that humans wear tight clothes. Applying the proposed method to people wearing other clothes will increase the error. The limitation of the method of this model is that in the current training, it cannot deal with the posture that is obviously different from the neutral posture and contains self-occlusion [20-23]. This can be solved by generating a larger training set (including more prominent poses).

The encoder-decoder mechanism is often used in the CNN model. CNN encoder can basically be regarded as the feature extraction network, while the decoder interprets the received information and enlarges the received image features to the size of the original image through the decoded part [16,24], which greatly facilitates the prediction of the part of the image segmentation area and simplifies many subsequent calculation processes.

Considering the calculation speed and accuracy of the above models, based on the Har DNet convolution neural network, a segmentation model of intestinal polyp image with a block convolution attention mechanism is proposed in this paper. The model uses Har DNet as the basic network structure, and the image processing results obtained after passing through the backbone network are augmented to varying degrees and input into the subsequent decoder. Firstly, the decoder uses the Convolutional Block Attention Module and RFB module [25] to process the image features, so that the model can improve the feature extraction ability of the network by simulating the human receptive field. At the same time, the input feature information is weighted through the block attention mechanism, combined with the low-level and high-level feature input combination representation, so that the overall accuracy and speed of feature extraction of the model are greatly improved.

In the paper "An Imperial Study on Ensemble of Segmentation Approvals" [29], it is proposed that many machine learning models cannot be continuously learned especially neural networks. When they receive new training tasks, they will forget what they have learned before. This phenomenon is often called catastrophic amnesia. One of the reasons for the assumption of catastrophic forgetting in neural networks is the change of input distribution in different tasks [12], for example, the lack of common factors or structures may cause the optimization method to converge to completely different results each time. This paper studies the catastrophic forgetting phenomenon that occurs when the model is learning the same task. The author tries to understand the optimization

process in depth by analyzing the interaction between samples during learning. The first goal is whether to continue compressing data sets based on some forgotten data, so as to improve the efficiency of data training without affecting the generalization accuracy. The second goal is to identify "important" samples, outliers and other special data by forgetting statistical data. The model used in this paper also has the defect of unsustainable learning. In the future improvement, we will consider the method of using statistical forgetting data to identify "important samples", combined with the attention mechanism module to extract special data for image processing.[13]

### III METHODS

#### A. HarDNet

Compared with the traditional CNN, Har DNet optimizes its sampling strategy and loss function based on L2-Net, reduces the inference time, and improves the accuracy of the results by using 1 convolution. Taking sampling in a batch as an example, Har DNet's sampling strategy selects the nearest negative sample by comparing the distance between the descriptors of the corresponding patch in the batch, so as to improve the inference speed of the network by reducing shortcuts. At the same time, Har DNet also optimizes its loss function based on this idea: the distance between matching pairs is less than the distance of unmatched patch time, which reduces unnecessary calculation and improves the inference speed of the network. Compared with the speed of the optimized network, the accuracy of the optimized network is better than that of the optimized network. Therefore, Har DNet is selected as the basic network structure of this model. Fig. 1 shows the network structure of Har DNet.

#### B. Decoder

The decoder generates more accurate mapping by combining low-resolution, high-level features, and high-resolution, low-level features, and refines deep features by cascade connection. However, this method has the following two disadvantages. First, the low-level features contribute less to the network performance than the high-level features. Secondly, the use of high-resolution low-level features makes the computation more complex.

According to the above two shortcomings, it is proposed to refine the deep-seated features to improve the representation ability and abandon the shallow features to reduce the computational complexity, that is, the cascaded partial decoder is used. By cascading, the original modules are connected across to improve the expression ability and reduce the computational complexity.

#### C. Convolution Block Attention Mechanism

The decoder, CBAM is also adopted to improve the computing power of the network. The visual attention mechanism of the human visual system can assist the brain in signal processing. The CBAM works similarly: the ultimate goal is to select the information that is more critical to the current task goal from a large amount of information.

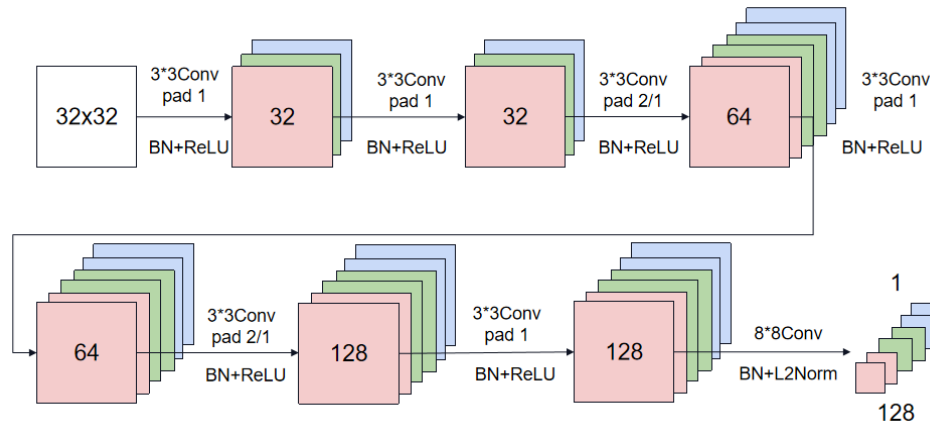


Fig. 1. HarDNet.

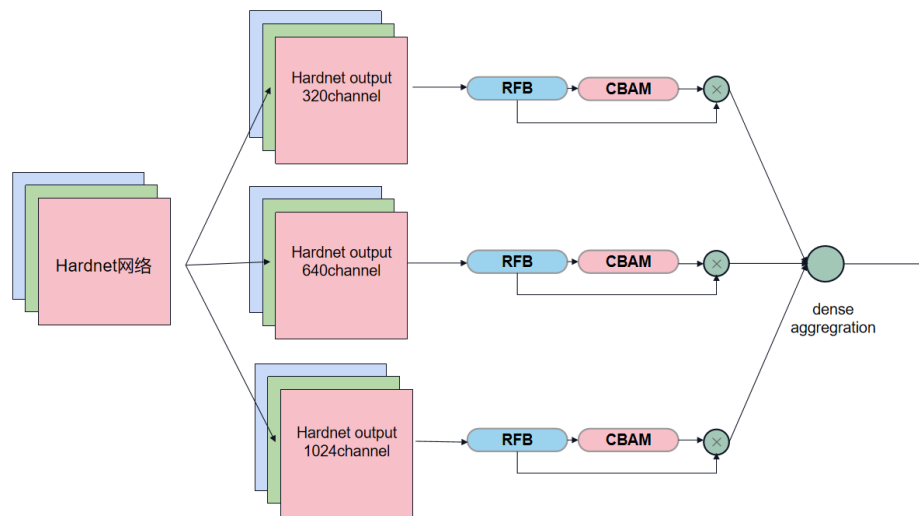


Fig. 2. Elaborate of convolutional block attention module.

The human visual attention mechanism receives the location of focal regions by quickly scanning all image information that needs to be processed by the brain and then pays more attention to these focal regions to obtain the detailed information needed by the brain. Through the judgment and selection of the focus of attention, the human eye pays more attention to the image information worth attention to rather than the non-focal, unimportant information. This attention mechanism of human vision greatly improves the information-processing efficiency of the human eye. At the same time, by investing more attention resources in the focus, the visual information obtained also has higher accuracy. Fig. 2 shows the action mechanism of the attention mechanism CBAM in the model.

The process of using the attention mechanism to improve the computing power of the network is similar to the human eye attention mechanism. Its core idea is to introduce specific weights into the input information in order to give priority to

the location of relevant information. This part of processing refers to the content of the Convolutional Block Attention Module (CBAM), and the input information is weighted through channel attention convolution and spatial attention convolution to enhance the useful features and suppress the useless features.

The channel attention model mainly focuses on the meaningful part of the input characteristic diagram, calculates the internal relationship between each channel, and finally uses the channel attention diagram to represent the internal relationship between each channel in the input characteristic diagram, and confirms the corresponding weighting of different channels to determine which channels should be ignored. The spatial attention model mainly focuses on the internal relationship of the feature map at the spatial level, that is, the weighting corresponding to different regions, which is complementary to the processing results of the channel attention model at the spatial level.

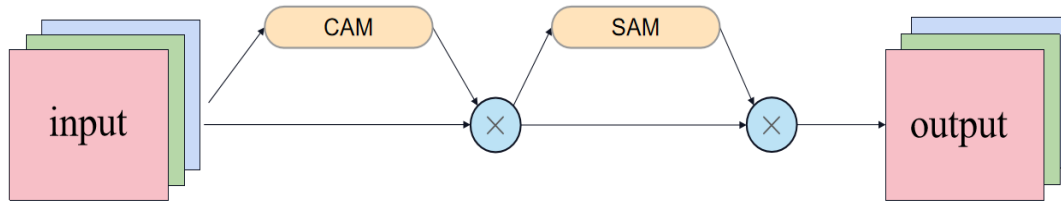


Fig. 3. Convolutional block attention module.

Fig. 3 shows the implementation mechanism of the channel attention mechanism in the model. It mainly processes the results obtained by the RFB model through a channel attention mechanism to generate the characteristic results of the channel attention mechanism and then takes it as the input characteristic diagram of the spatial attention mechanism to obtain the final results of the attention mechanism processing. The input features of RFB pass through the two pooling layers of Max pooling and average pooling respectively and then pass through MLP respectively. The output features are added based on element-wise and sigmoid activation to generate the final channel attention feature map. Then the feature map is multiplied by the input feature map, and the output feature map is used as the input feature of the spatial attention mechanism.

Similar operations are carried out in the part of the spatial attention mechanism. The feature map obtained through the channel attention mechanism passes through the two pooling layers of max pooling and average pooling, respectively. The two results are combined based on the channel, and then the feature map of the spatial attention mechanism is generated through the sigmoid operation. Finally, the feature map is multiplied with the input feature map to obtain the final

feature, which can be input into the subsequent calculation of the density aggregation part.

As Fig. 4 implies, the receptive field module network is composed of the convolutional layer with different expansion rates and convolution kernels of different sizes. By integrating the deep and shallow feature maps, the feature recognition effect can be equivalent to that of human vision. Through this method, the features are processed to achieve the effect of expanding the receptive field.

The decoder of this model enhances or suppresses the target features by adding a convolution block attention mechanism, which effectively improves the operation efficiency and feature expression ability of the model, and generally improves the performance of the model.

When processing the input image, rich feature representations is required. Combining these feature representations can improve the inference effect of category and location. Therefore, deep aggregation is used to enhance the network structure in order to better integrate information. The bottom stage is connected to the high level by jumping to fuse the size and resolution, to achieve the effect of combined expression of feature representation.

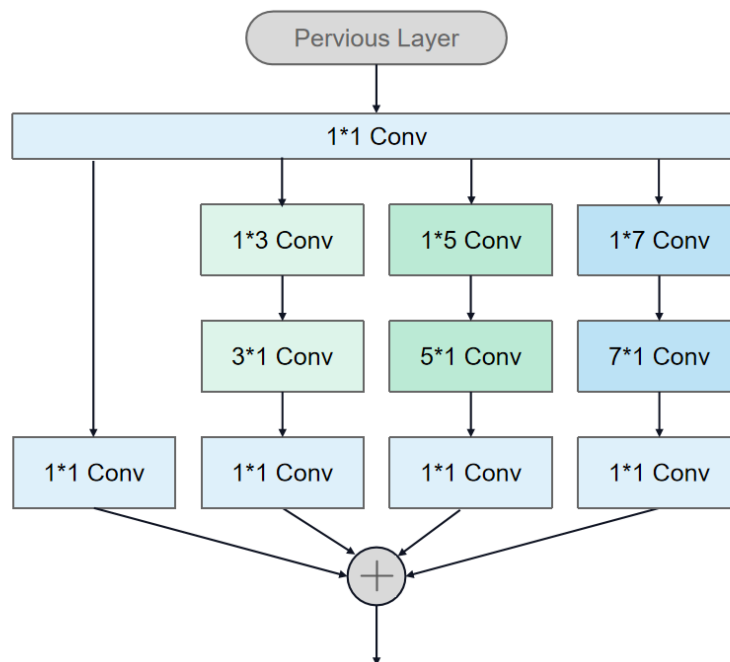


Fig. 4. Receptive field module network.

#### D. Algorithm Flow

Based on the HarDNet network, this model improved the processing speed and accuracy of the model by adding a

receptive field and block convolution attention mechanism. The overall algorithm flow is shown in Fig. 5.

---

**Algorithm 1:** Algorithm flow

---

```
initialization of the HarDNet network;  
while different size of the channel(320/640/1024) do  
    Processing through HarDNet module;  
    Processing through RFB module;  
    while input data processed by RFB do  
        Processing through CBAM module;  
        Multiplication CBAM and RFB modules' resylt;  
    end  
    Combine results through the dense aggregation module;  
end
```

---

Fig. 5. Algorithm

### IV EXPERIMENT

Based on the results of previous data processing in related polyp experiments, five polyp segmentation data sets were taken as the experimental data sets. At the same time, the network structure used in this experiment was compared with the previous network structure, and the reasoning speed and the accuracy of the results are compared.

#### A. Datasets

In this experiment, five polyp segmentation data sets mentioned earlier were used: Kvasir-SEG, CVC-ColonDB, EndoScene, ETIS-Larib Polyp DB, and CVC-ClinicDB. This section also compared the results obtained this time with the results obtained from several previous models: U-Net, U-Net++, ResUNet, SFA, and PraNet. Table I shows the experimental results of these models on CVC-T Data Sets.

#### B. Segmentation Results of Intestinal Polyp Image

The CVC-T data set mentioned earlier in the experiment is the largest publicly released gastrointestinal image data set, which includes images of pathological tissues and images of normal phenomena. According to the needs of this experiment, 1,000 pictures of intestinal polyps were selected for the experiment (polyps of different sizes, pathological positions, and different qualities), which can reflect the more

comprehensive learning ability of this model to a certain extent. At the same time, this experimental result was the calculation result of the above-mentioned evaluation index. Each column in the table below corresponds to an evaluation index respectively. Each behavior is different and used to compare the model. The model result of the last behavior experimental design, and n/a means that the result cannot be obtained.

In the above experiment, 512×512 training input size and 1e-2 learning rate were adopted. There were 100 epochs used to test the model. It can be seen that the reasoning speed with the attention mechanism is much faster than that of other models.

In addition, according to the above calculation results, the two evaluation indexes of mDice and mIoU are improved, in which the average intersection and union ratio is increased from 0.834 to 0.851, up to 0.17, and the accuracy and processing speed are improved. It shows that the experimental model can effectively segment the meat in the actual pathological picture. However, the SM index is 0.02 lower than that of the HarDNet model, because in the process of adding attention mechanism, the image is weighted and the processing weight of some pixels is reduced, which makes this experiment lack the similarity with the actual image.

TABLE 1. COMPARISON OF EXPERIMENTAL RESULTS OF DIFFERENT MODELS ON CVC-T DATA SETS

Models	mDice	mIoU	$F_{\beta}^{\omega}$	$S_m$	MAE	$E_m^{\max}$
U-Net	0.818	0.746	0.794	0.858	0.055	0.893
U-Net++	0.821	0.743	0.808	0.862	0.048	0.91
ResUNet-mod	0.791	n/a	n/a	n/a	n/a	n/a
ResUNet++	0.813	0.793	n/a	n/a	n/a	n/a
SFA	0.723	0.611	0.67	0.782	0.075	0.849
PraNet	0.898	0.84	0.885	0.915	0.03	0.948
HarDNet-MSEG	0.901	0.834	0.871	0.935	0.008	0.974
HarDNet-CBAM	0.902	0.836	0.881	0.933	0.03	0.978

TABLE 2. EXPERIMENTAL RESULTS ON DIFFERENT DATA SETS

Models	Clinicdb		ColonDB		ETIS		Kvasir	
	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU
U-Net	0.823	0.755	0.512	0.444	0.398	0.335	0.71	0.627
U-Net++	0.794	0.729	0.483	0.41	0.401	0.344	0.707	0.624
Res UNet-mod	0.779	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Res UNet++	0.796	0.796	n/a	n/a	n/a	n/a	n/a	n/a
SFA	0.700	0.607	0.469	0.347	0.297	0.217	0.467	0.329
Pra Net	0.899	0.849	0.709	0.640	0.628	0.567	0.871	0.797
Har DNet-MSEG	0.909	0.864	0.721	0.641	0.678	0.605	0.891	0.824
Har DNet-CBAM	0.926	0.879	0.756	0.677	0.735	0.661	0.897	0.841

On the basis of the above experiments, four other data sets are used to test the results of the model, namely, ETIS, CVC-ClinicDB, Kvasir-SEG, and CVC-ColonDB. Table II shows the quantitative results of these data sets.

The four data sets used in the experiment have different characteristics. CVC-ClinicDB and CVC-ColonDB are both small-scale data sets, so they are all used in the test experiments; In addition, ETIS is a data set for early diagnosis of colorectal cancer, containing 196 images that are also used in test experiments. In this experiment, the Endosece dataset is divided into three subsets (training, verification, and test) as the previous ones did. This experiment mainly adopted the kvasir subset as the test set.

It can be seen from Table II that the results of the model with attention mechanism under the two indicators of mDice and mIoU show higher accuracy and better performance.

C. Evaluation Index

In the training, this experiment employed 900 images in kvasir SEG and 550 images in CVC-ClinicDB, a total of 1450 images as training images. The 80% of the pictures in the five datasets mentioned above (Kvasir-SEG, CVC-ColonDB, EndoScene, ETIS-Larib Polyp DB, and CVC-ClinicDB) are randomly selected as the training set, and 10% for validation, 10% for the testing set. In addition, meanDice and meanIoU were used to evaluate the accuracy and correlation of image-cutting results. The calculation method of the index is as follows:

$$\text{MeanDice} = \frac{2 \bullet \text{tp}}{2 \bullet \text{tp} + \text{fp} + \text{fn}} \quad (1)$$

$$\text{mIoU} = \frac{\text{tp}}{\text{tp} + \text{fp} + \text{fn}} \quad (2)$$

In addition, in order to further evaluate the network model, the experiment also took into account other metrics for evaluation:  $F_{\beta}^{\omega}$ ,  $S_m$ , MAE,  $E_m^{\max}$  and FPS. Among them,  $F_{\beta}^{\omega}$  is used to modify the Dice process to avoid the deviation of the results when the samples are relatively balanced. MAE is used to evaluate the accuracy of the result, and the size is at the pixel level. Because the above two metrics are pixel-level evaluation indicators, they lack the relationship between macro and global structure and ignore the evaluation of the similarity between them. Therefore, another evaluation index  $S_m$  is used as the evaluation standard of the similarity between the prediction result and the actual image, so as to make up for the defect of only pixel-level evaluation. The alignment measure for enhancement is also used  $E_m^{\max}$  to evaluate the similarity between the image processing at the pixel level and the global image.

Fig. 6 is an exemplificative picture of the segmentation in the processing test data set. The masked part is the display part of the intestinal polyps after image segmentation, and the white part is the judgment position of the trained intestinal polyps. After comparison, we can roughly feel the approximate accuracy of the experimental results.

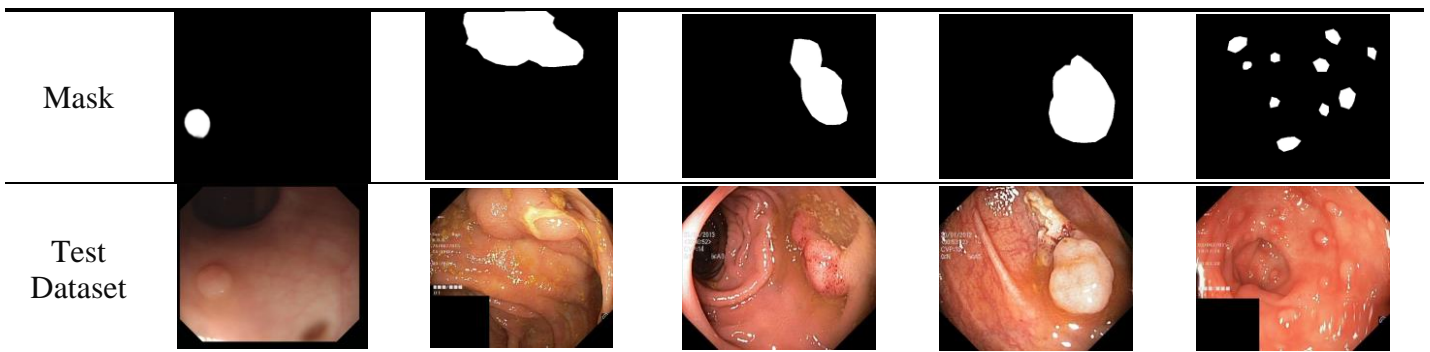


Fig. 6. Sample of model image segmentation results.

## V SUMMARY

In this paper, a segmentation model structure of intestinal polyp image based on attention mechanism was proposed. This model added Convolutional Block Attention Module (CBAM) and RFB module based on the HarDNet backbone to optimize the feature extraction ability. By enhancing or suppressing the target features in the computer operation process, the operation efficiency of the model and the expression ability of features were effectively improved. The fusion of the CBAM module and RFB module in the calculation details not only increases the receptive field of the model through RFB model processing but also makes the result of the final feature map more prominent through the attention mechanism, which greatly improves the ability of the model to extract image features and improves the processing speed and accuracy of the model.

The HarDNet-CBAM model has also obtained very good experimental results through experiments. The average precision and processing speed of four test data sets have been greatly improved, indicating that this model has been greatly improved in feature extraction. At the same time, the HarDNet-CBAM model has been tested in different data sets. Compared with other models, this model has higher accuracy and greater advantages in a variety of test sets. In a word, the HarDNet-CBAM model has excellent learning ability and universality in polyp image segmentation. At the same time, it also reflects its advantages in feature extraction performance and accuracy of results.

## VI FUNDING STATEMENT

The research is funded by the National Key R&D Program (2018YFB1307004); Supported by the National Natural Science Foundation of China (61873008).

## REFERENCES

- [1] Y. Cheng, et al. "Text Sentiment Orientation Analysis of Multi-Channels CNN and BiGRU Based on Attention Mechanism". *Journal of Computer Research and Development* vol. 12, pp. 2583-2595, 2020.
- [2] D. Xu, H. Li, L. X. Zhou, "Model of Automatic Identification of Diabetic Macular Edema via Convolutional Neural Networks UNet". *Recent Advances in Ophthalmology* vol. 4, pp. 357-361, 2020.
- [3] T. Zhou, et al. "U-Net and its applications in medical image segmentation: a review". *Journal of Image and Graphics*, vol. 9, pp. 2058-2077, 2021.
- [4] H. Zhang, et al. "Improved U-Net Models and Its Applications in Medical Image Segmentation: A Review". *Laser & Optoelectronics Progress* vol. 2, pp. 55-71, 2022.
- [5] Y. W. Zhou, et al. "Breast Mass Image Segmentation Algorithm Based on Improved Residual U-Net". *Journal of Southwest University of Science and Technology (Natural Science Edition)* vol. 2, pp. 68-74, 2021.
- [6] Y. Luo, "A Method for Segmentation Brain Tumor in MRI Based on ResUnet Embedded in GA", *Microcomputer Applications*, vol. 7, pp. 13-15, 2020.
- [7] Akbari, M., et al. (2018). "Polyp Segmentation in Colonoscopy Images Using Fully Convolutional Network." 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 69-72, 2018.
- [8] D. P. Fan, et al, "Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images". *IEEE Transactions on Medical Imaging*, vol. 39, pp. 2626-2637, 2020.
- [9] D. P. Fan, et al. "PraNet: Parallel Reverse Attention Network for Polyp Segmentation.". *MICCAI 2020: Medical Image Computing and Computer Assisted Intervention–MICCAI*, voi. *Medical Image Computing and Computer Assisted Intervention–MICCAI*, pp. 263-273, 2020.
- [10] P. Chao, et al. "HarDNet: A Low Memory Traffic Network". *CoRR*, pp. 00948, 2019.
- [11] C. H. Huang, et al., "HarDNet-MSEG: A Simple Encoder-Decoder Polyp Segmentation Neural Network that Achieves over 0.9 Mean Dice and 86 FPS." *ArXiv* pp. abs/2101.07172, 2021.
- [12] F. Zhang, et al., "Persistent Fault Attack in Practice." *IACR Trans. Cryptogr. Hardw. Embed. Syst.* pp. 172-195, 2020.
- [13] S. Engels, et al. "SPFA: SFA on Multiple Persistent Faults." 2020 Workshop on Fault Detection and Tolerance in Cryptography (FDTC), pp. 49-56, 2020.
- [14] L. D. Huynh, & N. Boutry, "A U-Net++ With Pre-Trained EfficientNet Backbone for Segmentation of Diseases and Artifacts in Endoscopy Images and Videos". *EndoCV@ISBI*, 2020.
- [15] D. Jha, et al. "DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation". 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS). Pp. 558-564, 2020.
- [16] Chen, L.C et al., "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation". *ArXiv* abs/1802.02611, 2018.
- [17] X. F. Yang, et al., "Road Detection and Centerline Extraction Via Deep Recurrent Convolutional Neural Network U-Net". *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, pp. 7209-7220, 2019.
- [18] Diakogiannis, F. I., (2019). "ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data". *ArXiv* abs/1904.00592.
- [19] D. Jha, , et al. "ResUNet++: An Advanced Architecture for Medical Image Segmentation". *IEEE International Symposium on Multimedia (ISM)*, 225-2255, 2019.
- [20] W. T. Zhu, Y. F. Huang, L. Zeng, "AnatomyNet: Deep Learning for Fast and Fully Automated Whole-volume Segmentation of Head and Neck Anatomy". *Medical Physics* 2, 2019.
- [21] C. E. He, H. J. Xu, & Wang, Z. (2020). "Automatic Segmentation Algorithm for Multimodal Magnetic Resonance-Based Brain Tumor Images". *Acta Optica Sinica* 6, 2020.
- [22] W. J. Wang, J. X. Chen, J. Zhao, et al. (2019). "Automated Segmentation of Pulmonary Lobes Using Coordination-Guided Deep Neural Networks[C]". 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), April 8-11, Venice, Italy. New York: IEEE Press, pp. 1353-1357, 2019.
- [23] Y. Li, et al., "Asymmetric GAN for Unpaired Image-to-Image Translation". *IEEE Transactions on Image Processing*, vol. 12, pp. 5881-5896, 2019.
- [24] L. C. Chen, et al. (2018). "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. " *ArXiv* abs/1802.02611 (2018): n. pag, 2018.
- [25] S. T. Liu, et al. (2018). "Receptive Field Block Net for Accurate and Fast Object Detection. " *ArXiv* abs/1711.07767: n. pag, 2018.
- [26] Oktay, Ozan et al. (2018). "Attention U-Net: Learning Where to Look for the Pancreas. " *ArXiv* abs/1804.03999: n. pag, 2018.
- [27] C. L. Guo, et al., "SA-UNet: Spatial Attention U-Net for Retinal Vessel Segmentation. " 2020 25th International Conference on Pattern Recognition (ICPR), pp. 1236-1242, 2021.
- [28] W. C. Zhang, C. Fu, Y. Zheng, F. Y. Zhang, Y. L. Zhao, "HSNet: A hybrid semantic network for polyp segmentation". *Computers in Biology and Medicine*. Volume 150. 2022.
- [29] Nanni, Loris, Alessandra Lumini, Andrea Loreggia, Alberto Formaggio, and Daniela Cuza "An Empirical Study on Ensemble of Segmentation Approaches" *Signals* 3, vol. 2, pp. 341-358, 2022.

# A Novel Hybrid DL Model for Printed Arabic Word Recognition based on GAN

Yazan M. Alwaqfi<sup>1</sup>, Mumtazimah Mohamad<sup>2\*</sup>, Ahmad T. Al-Taani<sup>3</sup>, Nazirah Abd Hamid<sup>4</sup>  
Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia<sup>1, 2, 4</sup>  
Faculty of Information Technology and Computer Sciences, Yarmouk University, Irbid, Jordan<sup>3</sup>

**Abstract**—The recognition of printed Arabic words remains an open area for research since Arabic is among the most complex languages. Prior research has shown that few efforts have been made to develop models of accurate Arabic recognition, as most of these models have faced the increasing complexity of the performance and lack of benchmark Arabic datasets. Meanwhile, Deep learning models, such as Convolutional Neural Networks (CNNs), have been shown to be beneficial in reducing the error rate and enhancing accuracy in Arabic character recognition systems. The reliability of these models increases with the depth of layers. Still, the essential condition for more layers is an extensive amount of data. Since CNN generates features by analysing large amounts of data, its performance is directly proportional to the volume of data, as DL models are considered data-hungry algorithms. Nevertheless, this technique suffers from poor generalisation ability and overfitting issues, which affect the Arabic recognition models' accuracy. These issues are due to the limited availability of Arabic databases in terms of accessibility and size, which led to a central problem facing the Arabic language nowadays. Therefore, the Arabic character recognition models still have gaps that need to be bridged. The Deep Learning techniques are also to be improved to increase the accuracy by manipulating the strength of technique in a neural network for handling the lack of datasets and the generalisation ability of the neural network in model building. To solve these problems, this study proposes a hybrid model for Arabic word recognition by adapting a deep convolutional neural network (DCNN) to work as a classifier based on a generative adversarial network (GAN) work as a data augmentation technique to develop a robust hybrid model for improving the accuracy and generalisation ability. Each proposed model is separately evaluated and compared with other state-of-the-art models. These models are tested on the Arabic printed text image dataset (APTII). The proposed hybrid deep learning model shows excellent performance regarding the accuracy, with a score of 99.76% compared to 94.81% for the proposed DCNN model on the APTII dataset. The proposed model indicates highly competitive performance and enhanced accuracy compared to the existing state-of-the-art Arabic printed word recognition models. The results demonstrate that the generalisation of networks and the handling of overfitting have also improved. This study output is comparable to other competitive models and contributes an enhanced Arabic recognition model to the body of knowledge.

**Keywords**—Deep learning; convolutional neural network; generative adversarial network; arabic recognition; image processing

## I. INTRODUCTION

Text recognition is among the fundamental computer science technologies, especially in pattern recognition, computer vision, and image processing. Furthermore, in the field of pattern recognition, text recognition is intended to compete against the human capacity to read written text in terms of speed and accuracy by correlating character codes (e.g., Unicode) with characters images (i.e., graphemes). The recognition process is defined as the process in which characters in images of a text are recognised and detected, which are then converted into data and coded that can be understood by the machine [1].

Arabic is known to be in a complex form of characters. Therefore, developing a novel and accurate Arabic printed word recognition model is still available for study. This is primarily due to the Arabic language characteristics, such as typing from right to left, diacritical marks, cursive characters and overlapping. Moreover, several Arabic letters contain dots, Hamza, Madda, and diacritics, which are markings placed above or below the Arabic letters; any misrecognition of these dots may result in the incorrect representation of the character and, hence, the entire word [2], [3]. Furthermore, multidimensional challenges (language-independent) are faced in developing systems for the Arabic language, which include the lack and absence of Arabic text databases, low-scanning resolution images, font variations, and the complex layout of scanned images. These limitations above contribute to the failure to apply the techniques developed for other types of languages to the Arabic language.

On the other hand, the term "deep learning" refers to a specific kind of machine learning based on an artificial neural network inspired by the structure and function of the human brain [4]. It was established to assist machine learning in achieving one of its primary intentions, which is artificial intelligence. It relies on imitating or mimicking the process in which the human brain processes data and extracting patterns to learn and perform intelligent decisions. Furthermore, deep learning is about learning several levels of representation and abstraction that aid in understanding data like images, sound, and text. Deep learning architectures are based on automated learning from features with no extraction or prior knowledge by using several layers to extract the final and needed information from raw data. Deep learning approaches have widespread use in various fields, including computer sciences, bioinformatics, image and pattern recognition, face and voice

\*Corresponding Author.

recognition, medical image analysis, drug discovery, and health risk evaluation [5]–[8].

Meanwhile, deep learning models such as CNNs have been shown to be beneficial in reducing the error rate and enhancing accuracy in Arabic printed word recognition systems [9]. In contrast, these models have several drawbacks when dealing with Arabic printed word recognition, including poor generalisation ability [2][10], overfitting [11][12] and low reliability [9]. These drawbacks are due to the reliability of Deep learning models, which increases with the depth of layers, but the essential condition for more layers is an extensive amount of data. Since CNN generates features by analysing large amounts of data, its performance is directly proportional to the volume of data, as DL models are considered data-hungry algorithms. The number of trainable parameters in deep learning models runs into millions; thus, appropriate training can only be executed by utilising a large amount of data, meaning that as the number of data increases, the model's performance becomes better behaved. Meanwhile, the lack of training data leads to overfitting, which will generate faulty or inaccurate predictions and decreased reliability. Furthermore, the accuracy rate for Arabic printed word recognition systems stated by the majority of studies using deep learning models has reached very high levels, even though the suggested solutions might not be scaled to other problems of a similar type, where most of these systems were tested on small datasets created privately for a particular task, each with its own evaluation protocols and metrics, making direct comparison and objective benchmarking impractical and not fair. Additionally, some proposed systems prevented other researchers from comparing their findings by making the tested dataset inaccessible to the public. Thus, these comparisons are considered to have low reliability. These issues are due to the limited availability of Arabic databases in terms of accessibility and size, which led to a central problem facing the Arabic language nowadays [13], [14]. Therefore, the availability of large and benchmark datasets is crucial for Arabic printed word recognition.

Overall, the accuracy of Arabic printed word recognition remains a problem in the field of pattern recognition, given that it is still at an infant stage for the Arabic language compared to Latin-based languages. These challenges are attributed to the fact that most Arabic recognition research continues to face the increasing complexity of the system's performance, where few efforts have been made to enhance models of recognising Arabic printed words. This depicts that there is still vast room for enhancements [15] [16].

Using novel deep learning techniques can revolutionise the interest and surge in the field of character recognition. Recently, GANs have been shown to be incredibly effective in various image-processing applications. Concurrently, GANs can be utilised in numerous fields, such as generating high-definition images from low-definition images, producing photo-realistic images of objects, and transforming images across domains [17].

Few studies related to Arabic recognition using GAN have been developed. The adoption of the Convolutional Neural Networks (CNNs) and Generative Adversarial Networks

(GANs) in the proposed research comes from their unintuitive generalisation behaviours and their ability to adapt to the new unseen data that arise from the same distribution as that used when the model was learned. In this paper, we investigate the potential of GANs algorithms in generating text to synthesise a result that could be used to estimate the prospect of the research topic. To build an Arabic printed word recognition model, the proposed methodology focused on enhancing the accuracy and reducing error rates for all possible reasons by covering the most critical phases of Arabic recognition (classification phase and training phase), and this was accomplished by utilising deep learning neural network models (DNN); two GAN models, and one DCNN model.

## II. RELATED WORK

Recent advances in deep learning algorithms, as well as the outstanding results obtained from CNN in image classification and prediction [6], paved the way for researchers to apply them in the classification and recognition of Arabic printed words [18].

In [19], a model is proposed for recognising offline Arabic printed documents to resolve the issue of segmentation for the Arabic text. A pipeline of three neural networks is proposed; the first network model predicted the Arabic words' font size, which is then employed in training the subsequent two models upon normalising the word of 18 points font size. Then words are segmented into characters by the second network model. Meanwhile, the Arabic characters are recognised by CNN by utilising the segmented characters as its input. The significant features are then automatically extracted. To enhance the model's generalisation ability and decrease the overfitting problem, the authors increased the training data size by applying data augmentation techniques. An accuracy rate of 94.3% is achieved when assessed on the APTI dataset.

A deep hybrid learning model is proposed in [20] to recognise printed Arabic text in a variety of font types and fonts that imitate Arabic handwritten scripts. Two bi-directional short long term memory networks (BDLSTM) and five convolution neural networks (CNNs) are employed. The proposed model functioned end-to-end and segmentation free. The proposed model was tested on the APTI database and reached an accuracy of 94.32%.

Recently, research on deep learning for Arabic text recognition has gained more attention, even though available Arabic databases remain limited in accessibility and size. In contrast, researchers have now begun to see artificial data as a safe and practical alternative to overcoming the obstacles to data access. Although generative models have received a lot of attention in the field of machine learning, they had a limited influence before the development of GANs, which provide an alternate data source for the current Arabic recognition systems that use machine learning, allowing the addition of artificial samples to training data to enhance model generalisation. Several GAN variants have been suggested after its advent in various fields, such as image processing and picture super-resolution. Despite this, only a few works based on GANs have been applied to the Arabic language [21]–[24].



According to [24], GANs are a merging technique to learn deep representation without much-annotated training data. The GAN model comprises a pair of deep neural nets aimed at training two adversarial networks by using a form of a min-max game. These adversarial networks include an expert known as a discriminator and a forger known as a generator. Additionally, the generative model receives random noise vectors as input and strives to generate an output (fake images) similar to genuine images (real images). The generator (forger) aims to generate forgeries realistic images that can trick the discriminator.

Meanwhile, the discriminator or expert strives to differentiate between fake and genuine images by classifying forged images generated by the generator as fake and genuine images from the original sample as real. On the other hand, the generator tries to reduce (minimise) its loss by optimising the objective function while the discriminator seeks to increase (maximise) it. A scalar likelihood that the input belongs to real data distribution is what it produces as an output [22], [25]–[27].

GANs are proposed to address the shortcomings of other deep learning algorithms. As mentioned early, the basic principle of GANs is the ability of the generator to generate as realistic samples as possible to deceive the discriminator. In contrast, the discriminator seeks to discern between fake and real examples. Through adversarial learning, the generator and discriminator both get better [27], [28]. Therefore, this adversarial process gives GANs a notable stand out among other generative algorithms. More precisely, GANs are superior to further deep learning algorithms because they can parallelise the generation, which is not achievable with other deep learning techniques like PixelCNN, and the generator design has few restrictions.

On the other hand, GANs are an effective subclass of generative models, which can generate entirely new valid data without requiring mathematical assumptions or the knowledge of explicit true data distribution. These benefits make GANs ideal for numerous astounding implementations that provide unexpected and unseen results; most of these implementations concern image processing [29], [30].

Furthermore, generative modelling is regarded as a strategy for data augmentation, also known as GAN-based data augmentation, and that refers to constructing artificial instances from a dataset that preserves comparable features to the original set. The main objective is to expand the dataset to address the overfitting issue and enhance the generalisation ability [22], [31].

In [22], a SentiGAN model is proposed to address the lack of dialectal Arabic datasets by augmenting low-size datasets and producing a variety of high-quality sentences for five different dialects of Arabic. Five generators and one discriminator are utilised in the dialectal Arabic generating process for a particular dialect, i.e., one generator for each dialect. The generator is used to produce new samples (text), and the discriminator assesses those samples (text), along with a dynamic update, to ensure that the process operates automatically and unsupervised. Before enriching the desired datasets, two metrics; novelty and diversity, are utilised to

check the consistency and the quality of the produced dialectal Arabic text. The MADAR dataset is used, and more sentences are generated than in the original dataset. Experimental results show that the created datasets are reliable and valuable.

The conditional deep convolutional generative adversarial network is the model suggested by the study in [23], it used for the generation of Arabic handwritten isolated characters. The suggested model consists of two CNNs, a discriminator and a generator, and competing to learn. The discriminator is a DCNN architecture composed of four convolutional layers, followed by an FC layer. Each convolutional layer involves batch normalisation and dropout with Leaky ReLU activation. The sequence of convolutions' high-level output is flattened, joined to the one-hot encoded conditional formation, and sent to a dense (FC) layer before being output. The generator is a generative algorithm that attempts to predict each pixel value of an image; the width and height of the input have to be increased to achieve the required shape. The first layer's output is modified and sent through three further layers of transposed/fractionally strode convolutions to predict a single-channelled 32×32 image. Batch normalisation with ReLU activation is performed on each transposed convolutional layer. The proposed model is trained on the AHCD dataset. Experimental results showed the effectiveness of the proposed model. Qualitative and quantitative data demonstrate that the produced samples are relatively equivalent to the genuine ones in terms of diversity and quality. Thus, the proposed model provides an easy substitute for the restricted Arabic handwritten characters database.

GANs have a wide range of types, such as BasicGAN, WGAN, DCGAN, VanillaGAN and BiGAN. These types share a discriminator and generator as standard GAN components, but their architectural designs vary. Consequently, different GANs generate various outputs. These types of GANs are used by [32] to generate different Arabic handwritten characters. The researchers split the proposed approach into two parts, one for generating an image and the other for evaluating the quality of the generated images by assessing how realistic and fake the generated images are. Fréchet inception distance (FID) and native-Arabic human evaluations are utilised. Each GAN is trained separately and generates different outputs due to the differences in their architectures. A total of 40,485 Arabic handwritten characters are generated. According to experimental findings, WGAN performed better in FID, with an accuracy of 96%. In contrast, DCGAN performs better in native-Arabic human evaluation, with an accuracy of 35%.

Furthermore, GANs have shown outstanding results in automatically synthesising high-resolution, realistic images from text representations. In [33], DF-GAN and AraBERT architectures are combined to generate images conditioned with Arabic text descriptions. The initial step is to translate the instruments from English to Arabic on CUB and Oxford-102 flowers datasets and recreate a new dataset that fits utilising the Arabic text-to-image generation operation by using the DeepL-Translator on text descriptions. Secondly, the AraBERT-generated sentence vector dimension was decreased to match the input shape of the FD-GAN. Thirdly, AraBERT is merged with DF-GAN by feeding the sentence embedding

vector to both the DF-GAN generator and discriminator. The proposed approach is evaluated with FID and IS. According to experimental results, the proposed model obtained an accuracy of 3.21% and 3.01% on SI and an accuracy of 60.96% and 65.45% on the FID score in the CUB and Oxford-102 datasets, respectively.

In [24], a novel GAN-based adaptive data augmentation model is proposed to recognise Arabic offline handwritten text. The proposed model is divided into two parts; in the first part, the authors employed an adaptive data augmentation method to generate a balanced dataset. Then, a GAN-based words synthesiser was trained in the new dataset to create images of Arabic words that look like they were written by humans, thereby boosting class diversity. The recurrent neural network library (RNNLIB) is used to develop the experimental models. The proposed model is evaluated on the IFN/ENIT and AHDB datasets and achieved an accuracy of 97.15% and 99.30%, respectively.

Based on the studies mentioned above, only a few of them are conducted on recognising the Arabic language by exploiting GANs to improve the recognition models and enhance accuracy rates, with generally satisfying results. Nevertheless, this is a worthwhile observation about the research in this field. Motivated by these research opportunities, the scarcity of studies and research motivated us to explore the strength of GAN to use it to recognise Arabic printed words to achieve promising results, as this study is considered among the first attempts in this field.

### III. MATERIALS AND METHOD

This section presents the proposed methodology of two deep learning-based models for Arabic printed word recognition; the DCNN model and the hybrid deep learning model based on GAN as a data augmentation technique. Also, it discusses the framework design of the proposed models.

#### A. Dataset

The training datasets are very crucial in the proposed approach since it is a straightforward way to get started with deep learning, and without having a good and sufficient training dataset, the system would never produce results with a high accuracy level. This study utilised the Arabic printed text image (APTI) dataset for testing and training. The APTI dataset is a collection of images of Arabic printed words. It was recently published by [34] for large-scale benchmarking of open-vocabulary, multi-font, multi-size and multi-style text recognition systems in Arabic. There are a total of 113284 different single words in the APTI dataset, each available in 10 diverse fonts: Andalus, Advertising Bold, Arabic Transparent, DecoType Thuluth, Tahoma, Traditional Arabic, Simplified Arabic, M Unicode Sara, and DecoType Naskh. It also comes with 10 different font sizes and four different styles: font sizes (6-18 and 24 points) and styles (plain, italic, bold, and a combination of italic and bold). For the present study, the dataset was split into training, validation and testing sets based on a ratio of 80%, 10%, and 10%, respectively.

#### B. Pre-processing Step

The success of classification for any DNN might be impacted by the pre-processing of the dataset images [35], affecting the final findings' recognition accuracy rate. In this work, both the DCNN model and the hybrid GAN-based model share the same pre-processing step, which involves dataset image resizing and image grey-scaling.

The training dataset images are comprised of various sizes. Therefore, this requires standardising the height and width of these images to cope with a dataset of images that could differ in size (height and width in pixels) throughout the stages of developing DNN layers. The size of the images should adjust before being used as input to the model. Therefore, images were resized to the size of 32×32 pixels.

On the other hand, the issue with images possessing multiple colour channels is that machine learning algorithms, or DNNs, have to function with three different data (R-G-B) values in order to extract the images' features and categorise them. This increases the computational complexity of the operation. DCNNs are responsible for transforming complex images into simpler ones for faster processing without missing pertinent features that are crucial to making an accurate prediction. Consequently, images are transformed into grayscale.

#### C. Proposed DCNN Model

The proposed DCNN model is based on two consequent convolutional layers, each possessing 64 filters, 3×3 kernel size and stride 1. Each layer is followed by a batch normalisation layer and a pooling layer of 2×2. Then the output of these layers is flattened by a flatten layer and squashed into two consequent fully connected dense layers, each possessing 256 neurons. Excluding the last layer, which used the softmax activation function, all other convolutional and dense layers in the proposed DCNN model employed the Leaky ReLU activation function. Lastly, the output of the two dense layers is keyed into the output layer – a fully connected dense layer that uses the softmax activation function to produce the results. The last layer comprises  $n$  neurons that represent the number of output classes. Fig. 1 depicts the overall architecture of the proposed DCNN model. For the optimisation process in the DCNN model, the Adam optimiser is adopted.

Moreover, the sparse categorical cross-entropy function is utilised for a loss function. To achieve the enhancement of the DCNN model, three significant modifications were made to the DCNN architecture. First, the DCCN model is modified by adding a batch normalisation layer after each max pooling layer in the model to catalyse and speed up training while utilising higher learning rates, which simplifies the learning process. Second, the Leaky ReLU activation function is adopted to replace the regular ReLU function in the dense and convolutional layers of the DCNN model to catalyse and faster the training duration, excluding the output layer, which utilises the softmax activation function. The final modification step is applying the sparse Categorical cross-entropy as a loss function to solve the problem of conversion long time.

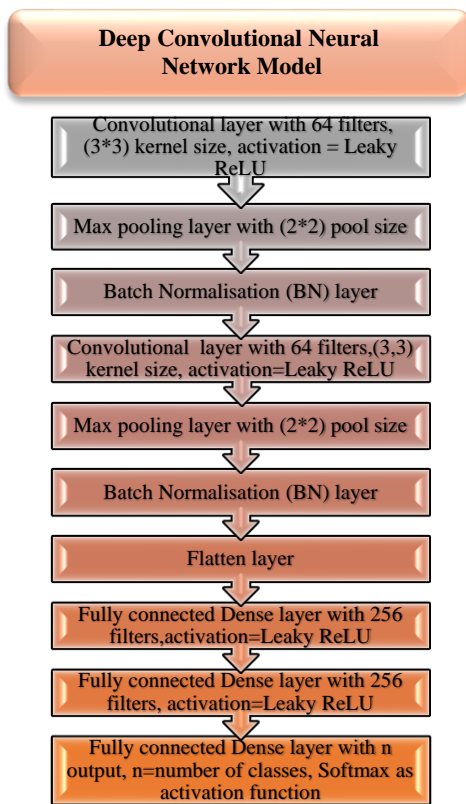


Fig. 1. DCNN architecture

The proposed DCNN model consists of ten layers. Fig. 2 shows the general framework flow of the adopted DCNN model.

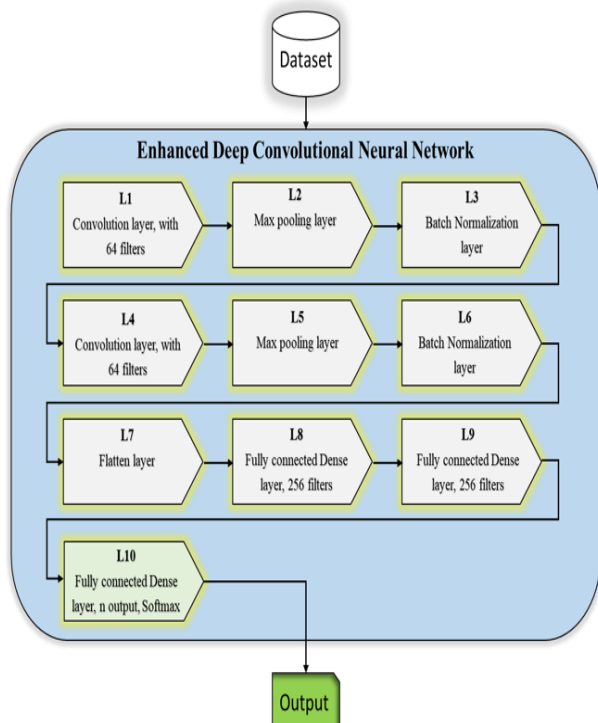


Fig. 2. DCNN model framework flow

As shown in Fig. 2, the proposed DCNN model layers are listed in Table I.

TABLE I. DCNN MODEL LAYERS

Layer no.	Layer Description
i	First layer is a 2D convolution layer (Conv2D). It uses 64 filters with a 3x3 kernel size. The training phase is catalysed and sped up by utilizing the Leaky ReLU as the activation function while preventing consistent predictions for negative input values.
ii	The second layer is a max polling layer with a 2x2 pool size. A pooling layer is a new layer added after the convolutional layer, particularly after nonlinearity, for instance, following applying Relu on the feature maps output by a convolutional layer. The dimensions of the feature maps can be reduced by using the pooling layers. Hence, the number of parameters required to learn and the amount of computation performed in the network are reduced.
iii	The third layer is the Batch Normalization layer. In each iteration, it normalizes the inputs by subtracting their mean and dividing by their standard deviation, and then it applies a scaling coefficient and a scale offset. After applying standardization, the resulting mini-batch has zero mean and unit variance.
iv	A 2D convolution layer (Conv2D) comprising a 3x3 kernel size and 64 filters makes up the Fourth layer. The layer's training was sped up by applying the Leaky ReLU as an activation function.
v	The fifth Layer is another max polling layer with a 2x2 pool size.
vi	The sixth layer is another Batch Normalisation layer.
vii	The seventh Layer is a Flatten layer, which is utilized to map the output of the previous Conv2D layer. After that, the Conv2D layer is flattened to be applied in the Dense layer as input. Some Neural Network implementations might not be able to map a spatial structure directly into a dense layer; thus, a flatten layer is necessary in between. Therefore, if the rank is higher, like with convolutions, the layer is implicitly flattened at first. So, the Flatten layer moves from convolution 2D to the Dense layer.
viii	The eighth layer is a fully connected dense layer using 256 filers and implementing Leaky Relu as an activation function.
ix	The ninth layer is a fully connected dense layer using 256 filers and implementing Leaky Relu as an activation function, which comprises a fully connected layer to transition from feature maps to an output prediction for the model.
x	The tenth layer is the output layer for the enhanced DCNN model. This layer uses the dense layer with the number of N+1 classes, and it implements the Softmax as an activation function.

Overfitting is one of the core difficulties that occur in multilayer NN learning, especially in deep neural networks. Overfitting generally happens once the DNN model performs effectively on the training dataset but fares badly on the testing dataset. Batch normalisation and dropout are two well-known approaches to overcoming this obstacle. Despite the fact that the two approaches share overlapping design fundamentals, multiple study findings have proven that they each offer particular advantages for enhancing deep learning.

However, several recent works have shown that the dropout technique is ineffective or even detrimental to CNNs training, which has proven that there are drawbacks to use in convolutional layers and sometimes its inefficiency. In

contrast, the training duration is typically increased upon applying dropout to a neural network, given that dropout increases the size of the network, which eliminates units during the training phase, and reduces the network capacity [11], [12].

On the other hand, adding batch normalisation in DNN architecture significantly improves the accuracy. Furthermore, batch normalisation substantially reduces training time and permits the utilisation of high learning rates. As a result, the training steps that are needed for network convergence are reduced [36].

According to [37], adding batch normalisation layers after the non-linear layers improves accuracy. Before introducing batch normalisation (BN), the network's training time was substantially dependent on the precise hyperparameters initialisation, including the adoption of modest learning rates, thereby increasing the time of training. During DNN training, each layer's input value distribution is influenced by the layers which came before it. Because this variation slows training, BN was developed to address it and accelerate the process of learning. The process of BN involves adjusting the unit value of every batch, and since batches are generated at random throughout training, additional noise is introduced into the training process. The noise works as a regulariser, and its effects are like that generated by dropout. Resultantly, dropout can be removed from the deep neural network.

According to [36], the addition of batch normalisation layers should be among the initial steps in optimising a CNN, not just for handling overfitting but also for its ability to improve accuracy significantly. To overcome the overfitting problem in the proposed enhanced DCNN Model, the batch normalisation layers are adopted and used after each convolutional layer in the network model.

#### D. Hybrid Model Based on GANs

This section presents the procedures applied to develop the proposed hybrid GAN-based model. The proposed hybrid GAN-based model combines the generative adversarial networks and the DCNN. The architecture of the proposed hybrid model consists of a GAN model that is used as a data augmentation technique and a DCNN model as a classifier employed for printed word recognition.

First, the GAN model is built, which comprises two deep neural networks; the discriminator network and the generator network. The generator generates new images, whereas the discriminator determines the originality of the generated images. The discriminator evaluates each produced image to decide if it is part of the training set or not. Nevertheless, the generator attempts to generate an additional fake instance in order to convince the discriminator that the image generated corresponds to the training set. During the training process, the discriminator and generator models compete with one another. While the generator strives to fool the discriminator, the discriminator attempts hard to avoid being deceived.

As a result, this competition between the two models encourages both to enhance their capabilities. The skills of these neural networks are sharpened as they undergo more training, thus increasing the ability of the generator to generate

real-like images which indistinguishable from that observed in the real world. Concurrently, the discriminator is more effective in detecting fake images. The accuracy of the final results will improve as the size of the training dataset increases. The training for the dataset will align with the level of the Arabic printed words. Fig. 3 shows the overall research framework for the proposed hybrid GAN-based model.

The second step entailed the classifier DCNN model development. The proposed DCNN is utilised to classify the images of Arabic printed words, which is the last step in the proposed hybrid GAN-based mode. The DCNN is employed, given that it is a typical multiclass classification NN. The classifier DCNN accepts the new dataset as input and captures various perspectives on features via the convolutional operation. The resulting feature maps produced from each convolution layer are passed to the following convolution layers with extra kernels. The reason for the pass is to extract a higher level of features of the input image. The pooling layer then summarises the feature map features generated by the convolution layers to reduce the dimensionality of the feature maps, hence minimising the amount of computation performed in the network. The classifier DCNN in the proposed hybrid GAN-based model shares the same architecture as the proposed DCNN model introduced previously. Fig. 1 shows the shared architecture used for both the stand-alone enhanced DCNN model and the classifier DCNN in the proposed hybrid GAN-based model. The DCNN model classifier uses the newly generated collection of images. This way, generative adversarial modelling is used as an alternative to data augmentation techniques in the proposed hybrid GAN-based model.

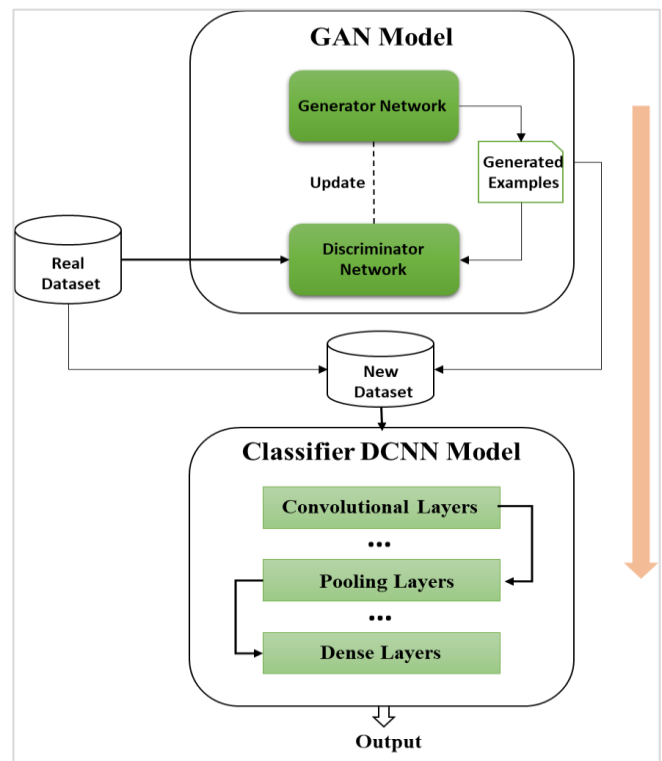


Fig. 3. Hybrid GAN-Based model design

1) *Generator neural network:* The proposed hybrid GAN-based model employs a generator neural network with ten layers. Two of the proposed enhancements for the proposed DCNN model are adopted in the generator neural network, the first one is the addition of adding batch normalisation layers, while the second one is adopting the Leaky ReLU as an activation function. The generator neural network layers are listed in Table II.

TABLE II. GENERATOR NEURAL NETWORK LAYERS

Layer no.	Layer Description
i	A sequential neural network of 10 layers is built with 8 layers as the hidden layers.
ii	The first layer is the input layer – a fully-connected dense layer with 32*32*256 filters, 32*32 for the width and height and 256 neurons.
iii	Second layer is a batch normalisation layer.
iv	Third layer is a reshape layer to convert the shape of data into two-dimensional representation.
v	Fourth layer is a transposed convolution layer with 256 filters and (3*3) kernel size.
vi	Fifth layer is another batch normalisation layer.
vii	Sixth layer is transposed convolution layer with 128 filters and (3*3) kernel size.
viii	Seventh layer is another batch normalisation layer.
ix	Eighth layer is another transposed convolution layer with 64 filters and (3*3) kernel size.
x	Ninth is the last batch normalisation layer in the network.
xi	The last layer is the output layer, which is another transposed convolution layer with (3*3) kernel size.
xii	The generator neural network is compiled using Adam Optimizer.
xiii	The loss in the generator neural network can be found with cross-entropy loss function.

2) *Discriminator neural network:* The discriminator neural network for the proposed hybrid GAN-based model consists of nine layers. For the additional enhancements, the batch normalisation layers are added after each convolution layer, and Leaky ReLU is adopted as an activation function. The discriminator neural network layers are listed in Table III.

TABLE III. DISCRIMINATOR NEURAL NETWORK LAYERS

Layer no.	Layer Description
i	A sequential neural network of 9-layers is built with 7 layers as the hidden layers.
ii	First layer is a convolution layer that has 64 filters and 3*3 kernel size with input shape 32*32*1 for height, width and one-color channel.
iii	Second layer is another convolution layer that has 128 filters and 3*3 kernel size
iv	Third layer is layer is another batch normalisation layer.
v	Fourth layer is another convolution layer with 256 filters and (3*3) kernel size.
vi	Fifth layer is another batch normalisation layer.
vii	Sixth layer is known as the flatten layer, this layer used to resize and flatten the data.
viii	Seventh layer is fully connected dense layer with 128 neurons (filters).
ix	Eighth layer (the last hidden layer) is another fully connected dense layer with 128 neurons (filters).
x	Ninth is the last batch normalisation layer in the

	network.
xi	The final layer is the output layer, a fully-connected dense layer with n+1 neurons (n = the number of classes).
xii	The discriminator neural network is compiled using Adam Optimiser.
xiii	The loss in the discriminator NN can be found with cross-entropy loss function.

3) *Deep CNN classifier:* The DCNN classifier in the proposed hybrid GAN-based model has the same architecture as the proposed DCNN model that introduced in section C. As mentioned previously, the DCNN classifier for the proposed hybrid GAN-based model consists of ten layers in total, and it has the same additional enhancements as the proposed DCNN model. The additional enhancements were the batch normalisation Layers, adopting the Leaky ReLU as the activation function and using sparse categorical cross-entropy as a loss function.

4) *Handling Overfitting for The Hybrid GAN-Based Model:* Since the proposed model consists of two parts; the GAN and the classifier DCNN part, two techniques were adopted to handle the possibility of overfitting in any of the two parts:

a) *Handling overfitting for the Generative Adversarial Part:* To avoid overfitting, generative and discriminative algorithms are used. The training dataset is fed to the discriminator; hence, the generator lacks any information on the training dataset while it creates new data instances. The generator is fed only from the discriminator, which decides whether each instance of data reviewed belongs to the actual training dataset or not. The proposed model is designed in a way that the generator never sees the genuine data; it must learn to create realistic information by receiving feedback from the discriminator. This process is called adversarial loss and works surprisingly well when implemented correctly.

b) *Handling overfitting for the Classifier Deep Convolutional Neural Network Part:* The significance of the proposed models in using GANs is to produce more realistic images to increase the dataset size for training the classifier deep CNN network. Among the techniques to handle overfitting is having more training datasets. Algorithms can be supported to detect signals effectively by training with more datasets. However, the technique would not be helpful if just noisier data were added to the training data. Thus, it should ensure that the newly generated dataset of images is clean and relevant in the GAN model. Actually, this technique works, such as data augmentation techniques, which can apply to increase the dataset's size artificially. Data augmentation refers to increasing the data size, which is increasing the number of images available in the dataset. As a result, in this way, generative adversarial modelling is utilised as an alternative to data augmentation techniques in the proposed hybrid GAN-based model. The use of generative adversarial modelling as a data augmentation technique offers a more domain-specific way to increase the number of training images with new high-quality generated images in domains with limited data. This approach enables the model to address the overfitting problem.

In addition to the previous solution, batch normalisation layers are adopted and added to the classifier DCCN part to tackle the overfitting problem in the proposed model. Since the DCNN classifier shares the same architecture as the DCNN model that proposed, a batch normalisation layer was added after each convolutional layer in the classifier DCNN to make an additional step in overfitting handling for the overall proposed hybrid GAN-based model.

#### IV. EXPERIMENTAL RESULTS AND DISCUSSION

Two models have been conducted in this study. The first model conducted is the DCNN model. Then the second model is a hybrid deep learning model based on GAN. In addition, this experimental design discusses the aim of the experiments, as well as how they may be attained. The two proposed models are evaluated using the Arabic printed text image database (APTI). The dataset is split into three sets (training, validation, and testing) using a ratio of 80%, 10%, and 10%, respectively. The implementation part of the proposed approach is a program written in Python using TensorFlow 2.0 and Keras.

##### A. DCNN Model

To evaluate the accuracy of the proposed DCNN model and to compare it with other approaches, the model is analysed through a number of evaluation scenarios using APTI datasets. The proposed DCNN is tested on the dataset before and after the additional enhancements. The results obtained from each experiment for the proposed DCNN model were compared before and after the enhancements.

Regarding the testing results, the proposed enhanced DCNN model achieved high (WRR) accuracy scores on the APTI dataset, which corresponded to an average overall test set and validation accuracy of 94.81% and 94.96%, respectively. On the other hand, the DCNN model without enhancements achieved an average accuracy of 93.67% and 94.17% for validation. The differences in results are shown in Fig. 5. The enhanced model on the APTI dataset starts learning with fast steps from the first epoch to the 20<sup>th</sup> epoch in a continuous increase in the accuracy until it attained an accuracy of 91%. Then, the training process started to slow down after the 20<sup>th</sup> epoch to the 40<sup>th</sup>. After that, the model reached stability in the training accuracy at epoch number 40. The model shows the accuracy results in ascending values and achieves the best accuracy in the 40 epochs with training accuracy.

From Fig. 4, the enhanced DCNN model learns better than the original DCNN model in the training phase. CNN models learn better if the training process is better; that means the enhancements of adding batch normalisation Layers, adopting the leaky ReLU as an activation function, and using sparse categorical cross-entropy as a loss function helped to make the model training process faster and more accurate in the Arabic printed words recognition and classification phase. Furthermore, the inclusion of batch normalisation layers facilitates the usage of higher learning rates during the training process. This explains why the enhanced DCNN model performed better in testing and validation results of the

enhanced DCNN model than the standard DCNN model without enhancements.

On the other hand, the proposed enhanced DCNN model for Arabic printed word recognition achieved high recognition accuracy (WRR) of 94.81%, outperforming other state-of-art models. Compared to the recognition accuracy obtained by the hybrid deep learning model based on CNN and BDLSTM to recognise printed Arabic text proposed by [15], where the enhanced DCNN model outperformed the hybrid CNN and BDLSTM Model by a small margin and achieved a recognition accuracy of 94.32%. In addition, the work by [14] proposed three neural networks pipeline model to recognise printed text images of the APTI dataset of words that contains over two-million-word samples, which uses data augmentation techniques and was able to score a recognition accuracy of 94.30%. The detailed comparison between the two-word recognition models is illustrated in Fig. 5.

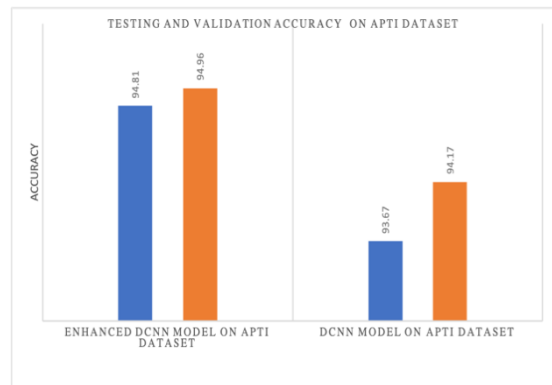


Fig. 4. Testing and Validation Accuracy for the DCNN Model with and without the enhancements

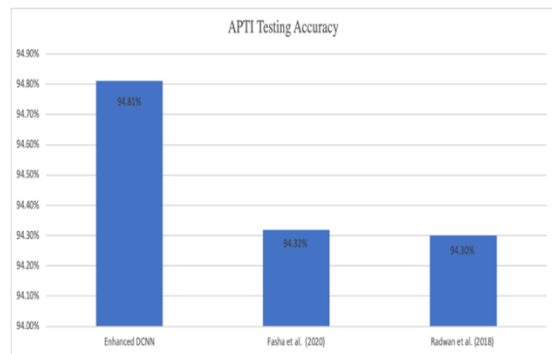


Fig. 5. Testing Results for the Enhanced DCNN against State-of-Art Models

##### B. Hybrid DL Model based On GAN

The performance of the proposed hybrid GAN-based model is evaluated by introducing the GAN phase into the proposed enhanced DCNN model to generate a new dataset of clean and relevant images. This occurs using GAN as a data Augmentation technique by artificially increasing the dataset size. The experimental testing for the proposed hybrid GAN-Based model is conducted in APTI datasets. Firstly, the results obtained from the proposed hybrid GAN-Based model were compared to those from the enhanced DCNN model to determine the impact of using the GAN phase, then compared

the proposed hybrid GAN-Based model with the state-of-the-art studies.

In terms of the testing results, very high accuracy scores (WRR) are achieved by the full proposed hybrid GAN-Based model with generative adversarial step on the APTI dataset. The overall test set accuracy is 99.76%, and 99.85% is achieved for the validation accuracy, while the corresponding values for the proposed enhanced DCNN model without the generative adversarial step as data augmentation technique was 94.81% and 94.96%, respectively. The discrepancies in results are presented in Fig. 6.

Fig. 6 illustrates that the proposed hybrid GAN-Based model achieved a higher success rate on the APTI dataset with higher testing accuracy results compared to the enhanced DCNN model. Thus, the hybrid GAN-based recognition model learnt better than the Enhanced DCNN model in the training phase.

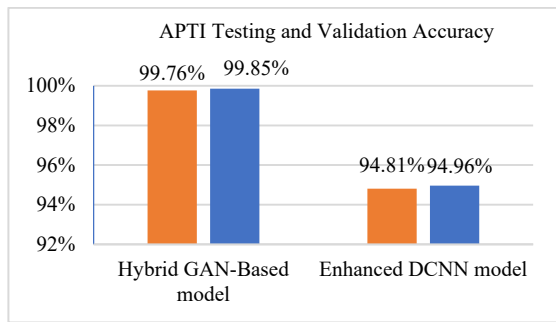


Fig. 6. Testing and Validation Accuracy for the proposed Hybrid GAN-Based model with the Enhanced DCNN model

Furthermore, as depicted in Fig. 7, the proposed hybrid GAN-Based model also achieved promising validation scores over the enhanced DCNN model. The proposed hybrid model that based on using the generative adversarial networks as a data augmentation technique achieved higher validation accuracy than the enhanced DCNN model. The overall validation and testing results of the hybrid GAN-Based model and the enhanced DCNN model on the APTI dataset are presented in Table IV below.

TABLE IV. APTI TESTING AND VALIDATION ACCURACY RESULTS

Implementation	Validation Accuracy %	Test Accuracy %
Hybrid GAN-Based model	99.85	99.76
Enhanced DCNN model	94.96	94.81

From the previous Fig. 6 and Table IV, it is notable that the hybrid GAN-based model that is based on using the GANs as a data augmentation technique outperformed the regular enhanced DCNN model in the APTI dataset, producing higher validation and test accuracy results. Adding GANs to the classifier DCNN model and using it as a data augmentation technique assisted in making the model training process more accurate in Arabic printed word recognition and classification.

As observed in the experiments, the hybrid GAN-based model commenced learning with fast steps from the first epoch to the 20<sup>th</sup> epoch with continuously increasing accuracy until a

value of approximately 93% is attained. Then, the training process started to slow down after the 20<sup>th</sup> epoch to the 40<sup>th</sup> one. After that, the model reached stability in the training accuracy at epoch number 40. The hybrid GAN-based model shows the result of accuracy in ascending values, achieving the best accuracy in the 40 epochs with training accuracy. The testing and validation accuracy measures utilise the word recognition rate (WRR) as a measurement unit on the APTI dataset.

On the other hand, the hybrid GAN-based model for Arabic printed word recognition achieved high recognition accuracy of 99.76%, thereby outperforming other the best state-of-the-art models. The proposed hybrid GAN-based model outperformed the recognition accuracy values obtained by the hybrid CNN and BDLSTM model proposed by [20] on 18 fonts and duplicate words across fonts types in the APTI dataset with a very good margin (99.76% vs. 94.32%), as well as the three NN pipeline model proposed by [19], that used traditional data augmentation techniques regarding recognition accuracy on the Printed Text Images of the APTI dataset and achieved recognition accuracy of 94.30%. The detailed comparison between the printed word recognition models is illustrated in Fig. 7.

The experimental results for the proposed hybrid GAN-based model demonstrated that the proposed model achieved very high accuracy (WRR) scores on the APTI dataset. Specifically, the proposed hybrid GAN-based model achieved average overall testing and validation accuracy of 99.76% and 99.85%, respectively. Meanwhile, the proposed enhanced DCNN model, without using any additional data augmentation techniques, achieved an average overall test set accuracy of 94.81% and 94.96% for the validation accuracy. The proposed hybrid GAN-based benefits from all the additional enhancements that are added to the enhanced DCNN model, batch normalisation layers were added to the proposed hybrid GAN-based model networks after each convolutional layer. The leaky ReLU was also utilised as an activation function, and the sparse categorical cross-entropy was employed in the classifier DCNN part of the proposed hybrid GAN-based model as the loss function.

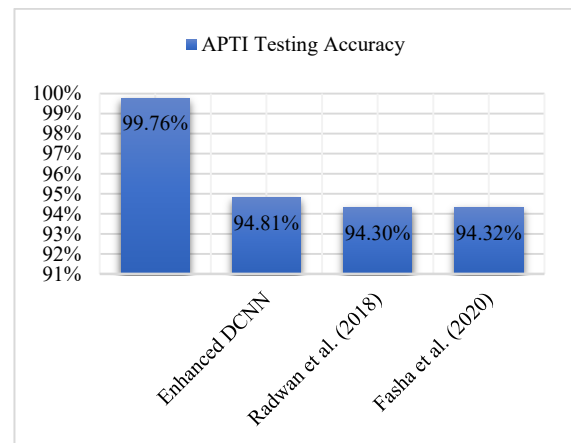


Fig. 7. APTI Testing Results of the proposed Hybrid GAN-Based model and other state-of-the-art models

From an analytical point of view, Fig. 6 and Fig. 7 depict the highly competitive testing accuracy results achieved by the proposed hybrid GAN-Based model, which outperformed other state-of-the-art models. The result proves that the proposed hybrid GAN-Based model could learn more and better during the training phase than the other machine learning-based models. The bigger training data generated from the generative adversarial model assisted the proposed hybrid GAN-based model in scoring higher recognition accuracy. This explains why the proposed hybrid GAN-based model outperformed other CNN models in the state-of-the-art models for Arabic printed word recognition on the APTI dataset.

## V. CONCLUSION

This study is considered one of the first to successfully develop a novel framework model based on DL techniques using DCNN and GAN to recognise Arabic words. The proposed model included two kinds of deep neural networks; GAN and DCNN, to enhance the training and classification phase of the printed word recognition model. The generative adversarial neural networks were designed to work as a data augmentation technique that is applied to artificially increase the size of the dataset by generating a new set of images and adding it to the original dataset to tackle the lack of Arabic datasets, the more extensive dataset will enhance the training accuracy for the Arabic text recognition classifier DCNN model. Additionally, solutions for the overfitting problem were introduced, such as having more data to train by employing the GAN as a data augmentation technique. More data training can assist models in avoiding overfitting issues. The experimental results showed the strength of the proposed approach of using the GAN model as a data augmentation technique.

The impressive testing accuracy results for printed word recognition prove that generative adversarial models can be used as a very good alternative data augmentation technique that can be applied to artificially increase the size of the dataset by new high-quality and real-like images indistinguishable from the real dataset images. Such new images can be crucial in boosting the training accuracy for the DCNN models. The results of the proposed model are promising, where the accuracy of the proposed model was 99.76%. These results prove the potential of the proposed model to develop an Arabic language recognition field. In future, we will test the proposed model on several languages to evaluate the proposed hybrid GAN-based model in terms of generalisation and token independency.

## REFERENCES

- [1] A. Chhabra, "Deep Learning Based Real Time Devanagari Character Recognition," 2019.
- [2] G. et al. Ahmed, R., "Novel Deep Convolutional Neural Network-Based Contextual Recognition of Arabic Handwritten Scripts," *Entropy*, vol. 23, no. 3, pp. 4–6, 2021, doi: 10.3390/e23030340.
- [3] R. Maalej and M. Kherallah, "Convolutional Neural Network and BLSTM for Offline Arabic Handwriting Recognition," *ACIT 2018 - 19th International Arab Conference on Information Technology*, pp. 1–6, 2018, doi: 10.1109/ACIT.2018.8672667.
- [4] A. Labach, H. Salehinejad, and S. Valaee, "Survey of Dropout Methods for Deep Neural Networks," 2019, [Online]. Available: <http://arxiv.org/abs/1904.13310>
- [5] H. Q. Ghadhban and M. Othman, "Survey of Offline Arabic Handwriting Word Recognition Survey of Offline Arabic Handwriting Word Recognition," no. February, 2020, doi: 10.1007/978-3-030-36056-6.
- [6] M. Makhtar, R. Rosly, M. K. Awang, M. Mohamad, and A. H. Zakaria, "A multi-classifier method based deep learning approach for breast cancer," *International Journal of Engineering Trends and Technology*, no. 1, pp. 102–107, Aug. 2020, doi: 10.14445/22315381/CAT13P217.
- [7] N. Altwajry and I. Al-Turaiki, "Arabic handwriting recognition system using convolutional neural network," *Neural Comput Appl*, vol. 33, no. 7, pp. 2249–2261, 2020, doi: 10.1007/s00521-020-05070-8.
- [8] Oueslati, "A Review of Sentiment Analysis Research in Arabic Language," *Future Generation Computer Systems* 112 (2020): 408-430, 2020.
- [9] L. F. Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al Dujaili, Ye Duan, Omran Al Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al Amidie, "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.
- [10] M. Jain, "Unconstrained Arabic & Urdu Text Recognition using Deep CNN-RNN Hybrid Networks," no. July, 2018.
- [11] Cai, Shaofeng, et al., "Efficient and Effective Dropout for Deep Convolutional Neural Networks," *arXiv preprint arXiv:1904.03392*, pp. 1–12, 2019.
- [12] Zeng, Yuyuan, et al. "Correlation-based structural dropout for convolutional neural networks," *Pattern Recognit*, vol. 120, p. 108117, 2021, doi: 10.1016/j.patcog.2021.108117.
- [13] T. Meriem, M. Asma, B. Rahima, K. Said, and B. Belgacem, "Recognition of Printed Arabic Characters".
- [14] I. Saleh Al-Sheikh, M. Mohd, and L. Warlina, "A Review of Arabic Text Recognition Dataset," *Asia-Pacific Journal of Information Technology and Multimedia*, vol. 09, no. 01, pp. 69–81, 2020, doi: 10.17576/apjtitm-2020-0901-06.
- [15] Darwish et al., *Arabic Offline Character Recognition Model Using Non-dominated Rank Sorting Genetic Algorithm*, vol. 1261 AISC, no. October. Springer International Publishing, 2021. doi: 10.1007/978-3-030-58669-0\_20.
- [16] R. Mohd, M., Qamar, F., Al-Sheikh, I., & Salah, "Quranic optical text recognition using deep learning models," *IEEE Access*, vol. 9, pp. 38318–38330, 2021, doi: 10.1109/ACCESS.2021.3064019.
- [17] Creswell, "Generative Adversarial Networks : An Overview," no. April, pp. 1–14, 2018.
- [18] A. Bhardwaj, "An Accurate Deep-Learning Model for Handwritten Devanagari Character Recognition," vol. 7, no. 2, pp. 1317–1328, 2022.
- [19] M. A. Radwan, M. I. Khalil, and H. M. Abbas, "Neural Networks Pipeline for Offline Machine Printed Arabic OCR," *Neural Process Lett*, vol. 48, no. 2, pp. 769–787, 2018, doi: 10.1007/s11063-017-9727-y.
- [20] J. A. Mohammad Fasha, Bassam Hammo, Nadim Obeid, "A hybrid deep learning model for arabic text recognition," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 8, pp. 122–130, 2020, doi: 10.14569/IJACSA.2020.0110816.
- [21] Y. M. Alwaqfi, M. Mohamad, and A. T. Al-Taani, "Generative Adversarial Network for an Improved Arabic Handwritten Characters Recognition," *International Journal of Advances in Soft Computing and its Applications*, vol. 14, no. 1, 2022, doi: 10.15849/IJASCA.220328.12.
- [22] X. A. Carrasco, A. Elnagar, and M. Lataifeh, "A Generative Adversarial Network for Data Augmentation: The Case of Arabic Regional Dialects," *Procedia CIRP*, vol. 189, pp. 92–99, 2021, doi: 10.1016/j.procs.2021.05.072.
- [23] S. M. S. Ismail B. Mustapha, Shafaatunnur Hasan, Hatem Nabus, "Conditional Deep Convolutional Generative Adversarial Networks for Isolated Handwritten Arabic Character Generation," *Arab J Sci Eng*, no. 0123456789, 2021, doi: 10.1007/s13369-021-05796-0.



- [24] Z. Eltay, et al. "Generative adversarial network based adaptive data augmentation for handwritten Arabic text recognition," *PeerJ Comput Sci*, vol. 8, pp. 1–22, Jan. 2022, doi: 10.7717/PEERJ-CS.861.
- [25] S. Barua, S. M. Erfani, and J. Bailey, "FCC-GAN: A Fully Connected and Convolutional Net Architecture for GANs," no. ii, 2019, [Online]. Available: <http://arxiv.org/abs/1905.02417>
- [26] A. P. Behera, S. Godage, S. Verma, and M. Kumar, "Regularized Deep Convolutional Generative Adversarial Network," *Communications in Computer and Information Science*, vol. 1378 CCIS, no. September, pp. 452–464, 2021, doi: 10.1007/978-981-16-1103-2\_38.
- [27] S. Khamekhem Jemni, M. A. Souibgui, Y. Kessentini, and A. Fornés, "Enhance to read better: A Multi-Task Adversarial Network for Handwritten Document Image Enhancement," *Pattern Recognit*, vol. 123, 2022, doi: 10.1016/j.patcog.2021.108370.
- [28] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications," *IEEE Trans Knowl Data Eng*, vol. 14, no. 8, pp. 1–28, 2021, doi: 10.1109/TKDE.2021.3130191.
- [29] T. Iqbal and S. Qureshi, "The Survey: Text Generation Models in Deep Learning.," *Journal of King Saud University - Computer and Information Sciences*, 2020, doi: 10.1016/j.jksuci.2020.04.001.
- [30] L. Jose, S. Liu, C. Russo, A. Nadort, and A. di Ieva, "Generative Adversarial Networks in Digital Pathology and Histopathological Image Processing : A Review," *J Pathol Inform*, vol. 12, no. 1, p. 43, 2021, doi: 10.4103/jpi.jpi.
- [31] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J Big Data*, 2019, doi: 10.1186/s40537-019-0197-0.
- [32] T. Alkhodidi et al., "GEAC: Generating and Evaluating Handwritten Arabic Characters Using Generative Adversarial Networks," *Proceedings of 2nd IEEE International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2021*, pp. 228–233, 2021, doi: 10.1109/ICCIKE51210.2021.9410746.
- [33] M. Bahani and K. Maalmi, "Fusion of AraBERT and DF-GAN for Arabic Text to Image Generation Fusion of AraBERT and DF-GAN for Arabic," pp. 0–16, 2022.
- [34] F. Slimane, R. Ingold, S. Kanoun, A. M. Alimi, and J. Hennebert, "A new Arabic printed text image database and evaluation protocols," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, 2009*, pp. 946–950. doi: 10.1109/ICDAR.2009.155.
- [35] Y. M. Alwaqfi and M. Mohamad, "A review of Arabic optical character recognition techniques & performance," *International Journal of Engineering Trends and Technology*, no. 1, pp. 44–51, 2020, doi: 10.14445/22315381/CATI1P208.
- [36] C. Garbin, X. Zhu, and O. Marques, "Dropout vs . batch normalization : an empirical study of their impact to deep learning," pp. 12777–12815, 2020.
- [37] C. S. Sergey Ioffe, "Australian Literary Journalism and 'Missing Voices': How Helen Garner finally resolves this recurring ethical tension," *Journalism Practice*, vol. 10, no. 6, pp. 730–743, 2015, doi: 10.1080/17512786.2015.1058180.

# Quantum Cryptography Experiment using Optical Devices

Nur Shahirah Binti Azahari, Nur Ziadah Binti Harun

Department of Information Security and Web Technology-Faculty of Computer Science and Information Technology,  
University Tun Hussein Onn Malaysia  
86400 Parit Raja, Johor, Malaysia

**Abstract**—The study of quantum cryptography is one of the great interest. A straightforward and reliable quantum experiment is provided in this paper. A half wave plate in linearly polarized light makes up a simplified polarization rotator. The polarization rotates twice as much as the half wave plate's fast axis' angle with the polarization plane when the half wave plate is rotated. Here, an experiment of message sharing is conducted to demonstrate quantum communication between parties. The unitary transformation is performed step by step using half-wave plates represented by the Mueller matrix. A simulation created using Python programming has been used to test the proposed protocol's implementation. Python was chosen because it can mathematically imitate the quantum state of superposition.

**Keywords**—Half-wave plate; polarizer; photon beam splitter; Stoke Vector

## I. INTRODUCTION

Numerous organizations, including those in the industrial and academic areas, have shown a great deal of interest in quantum cryptography. Numerous productive research projects over the past few years have demonstrated a great advancement in optical equipment development and quantum cryptography [1]. The term "flying quantum bits" is frequently used to describe photons, emphasizing both their quantum nature and their capacity to transport quantum information over long distances [2]. Light can be thought of as a transverse electromagnetic wave, as has already been established. So far, it only discussed light that is linearly polarized or plane-polarized, meaning that the electric field's orientation is constant despite changes in its amplitude and sign over time.

An optical filter known as a polarizer or polarizer is used to block light waves of certain polarizations while allowing light waves of others to flow through. A polarizer is also an optical device that uses natural light as its input and produces some kind of polarized light as its output [3]. This means that it may convert an undefined or mixed polarization light beam into polarized light by filtering it through a well-defined polarization beam. For instance, consider the superposition of two equal-amplitude, incoherent, orthogonal p-states as a possible representation of unpolarized light. A linear polarizer is a device that separates these two elements, discarding one and keeping the other [4]. In fact, the BB84 protocol, the very first quantum key distribution approach, was put forth using the polarization of photons.

Normally, light is free to oscillate at an angle to its path in any direction. When light passes through a polarizing filter, it becomes polarized. But how can be telling experimentally whether a device is a linear polarizer or not? Only light in a p-states will be transmitted if light is incident on a perfect linear polarizer. The orientation of that p-state will be parallel to the transmission axis of the polarizer, which is a specific direction. Only the part of the optical field that is parallel to the transmitting axis will effectively pass through the device unchanged. Due to the full symmetry of unpolarized light, the reading from the detector will not change if the polarizer is rotated about the z-axis. Based on the output beam's intensity, the polarizer determines the polarization states.

Now take a look at a type of optical components called retarders or also known as wave plate, which are used to modify the polarization of incident waves. Half-wave plates ( $\lambda/2$  plates) are the most common types of waveplates. Half wave plate is described as having  $\frac{\theta}{2}$  phase shift [5]. Linearly polarized light can be rotated by half wave retarders to an angle that is double the angle between the fast axis of the retarder and the plane of polarization. The polarization rotates  $90^\circ$  when the fast axis of a half wave plate is positioned at  $45^\circ$  from the polarization plane.

Continuously adjusting the energy with a half-wave plate and polarizer is another effective strategy. The polarizer can select the polarization direction of light, whereas the half-wave plate can rotate the polarization direction of light, and the two together may provide continuous light energy adjustment. The half-wave plate may continually change the polarization orientation. Assume that the slow axis is at an angle,  $\theta$  to the polarisation direction. The phase of the fast axis is incremented by one after passing through the polarizer, and the polarisation direction is rotated by an angle,  $\theta$ . Light passes into the polarizer through the half-wave plate. The transmitted energy for the polarizer is determined by Malus' law (Eq. 1).

The optical beam splitter is a crucial component in determining the statistical properties of light. The beam splitter has been used in many different areas of optics. For instance, in the field of quantum information, the beam splitter is fundamental for teleportation, bell measurements, entanglement, and basic research on photons. A beam splitter is a tool that splits an input beam into two beams travelling in different directions [6].

The rest of this paper is organized as follows: Section II reviews the related works of quantum cryptography that use optical devices. Section III presents the experiment environment. Section IV describes the principle of operation with schematic. Section V discusses hardware implementation. The experiment setup is discussed in Section VI. Section VII describes the experiment simulation. Pseudo code explained in Section VIII. Finally, Section IX concludes this paper.

In this study, presented a straightforward method for polarization alignment in which a half-wave plate is used to rotate the polarization of incoming light with respect to the fiber axis.

## II. RELATED WORK

In the last few years, several quantum experiment protocols using optical devices have been proposed. The main objective of those experiments was to ensure that the messages are delivered well among the parties involved. This section reviews several quantum experiments that use optical devices such as polarizer and half-wave plate for stimulation how the sharing message operation happens in quantum network, discussing their basic concepts, and describing their advantages and drawbacks. Fig. 1 shows the example implementation of polarizer, which when light passes through a polarizing filter, and it becomes polarized.

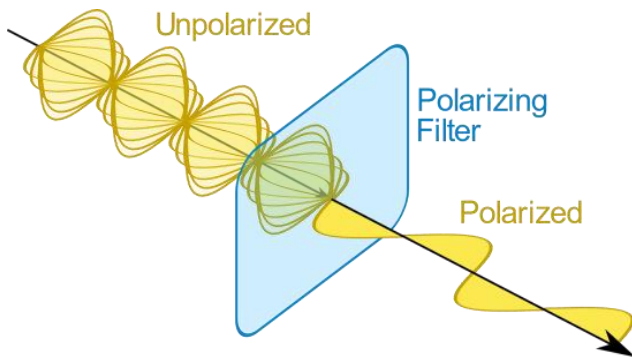


Fig. 1. Polarisation of light [7]

According to prior research, the quantum gate offers a number of advantages, including being simple to operate because parties do not always share information about the operators, requiring less precision than arbitrary rotational devices like half-wave plates (HWP) which enables precise measuring [8]. In 2019, Harun *et al.* proposed Hybrid M-ary in a braided single stage (HMBSS) [9]. The unitary transformation is performed step by step utilizing half-wave plates represented by the Mueller matrix. Five HWP was implemented in this protocol. For the purpose of authentication, the half wave plate's update angle or rotation is updated every 8 bits. The level of security has been raised at each stage due to the quick polarization changes, although it takes longer to send the information. The time required for message transmission and half-wave plate rotation is represented by the total transmission time. One can compute the transmission time as follows [9]:

$$\text{Transmission Time} = T_{msg} + T_{HWP} \quad (1)$$

where  $T_{msg}$  is the message transfer time and  $T_{HWP}$  is the half wave plate's time to update its angle during the encryption phase. The more the number of half wave plate implemented, the more time needed to encode each bit [10]. For example, the IV three-stage technique takes a long time because a single bit must exchange 3 times over the channel when using the three-stage protocol. The IV three-stage protocol requires Alice and Bob to use seven HWP for the polarization of bits in order to ensure the message's confidentiality, which lengthens the time it takes to encode each bit [11].

In 2021, with the use of single qubit unitary operations, Kang *et al.* have created a quantum message authentication mechanism that ensures the authenticity and integrity of the original message [12]. This protocol is made up of two parts: quantum encryption and a consistency check. Linear combinations of wave plates are used to implement the quantum encryption component. Assume that the secret key sequences have already been disclosed to Alice and Bob. This protocol implemented quantum encryption using half wave plates on Alice's side for the message authentication step.

Riggs *et al.* proposed Multi-Wavelength Quantum Key Distribution Emulation [13]. Multi-wavelength QKD can enable ternary and quaternary data transfer while removing the requirement to share a portion of the final key for eavesdropper detection. The half-wave plates were rotated before the laser pulse is sent in order to polarize the laser and measure it appropriately. In order for a half-wave plate to function, linearly polarized light must be rotated for the resulting polarization is twice as large as the angle between the optical axis and the incident polarized light [13]. In data transmission phase, through Alice's half-wave plate and Bob's half-wave plate, the laser beam is polarized. Bob's half-wave plate serves as the beam's measurement. At this point, the laser beam should only be vertically or horizontally polarized if Alice and Bob employed matching bases. If the resulting polarization is diagonal, the horizontally polarized light is transmitted while the vertically polarized light is reflected at the polarizing photon beam splitter cube. Just one of the detectors should experience the full intensity of the beam and light up if Alice and Bob employ matching bases.

## III. EXPERIMENT ENVIRONMENT

In order to fully examine the performance of the proposed method, numerous parameter values must be determined. The quantum network simulation photon generation, encoding, decoding stage and schematic diagram utilized in this paper are described in this section.

### A. Quantum Network Simulation

Quantum networks including hardware components such as optical devices are costly to deploy since expert configuration is required and real-time experiments take a long time to complete. As a result, quantum communication protocols are tested and evaluated using simulation.

Because it can describe quantum states in mathematical form, the Python software was used to simulate the function of hardware components in a quantum communication

environment. The Python software is also used to simulate the impact of light polarization in the implementation.

### B. Photon Generation

A sender turns the original message into a binary string, which is then converted into a qubit via photon polarization. Multiple photons were created for each encoded bit [14], referring to one bit of information for the production of multiphoton, let's say three photons for each bit. The operation begins with the laser generating a laser beam, which is then passed over the quantum channel and transmitted through a beam splitter, which splits the beam into two channels, one with a 0° polarizer and the other with a 90° polarizer. The mirror then assembles the laser beam before combining it with a beam combiner.

### C. Encoding and Decoding

For encoding stage, the beam is directed towards the HWP at this point in order to modify the polarization of the beam using a secret polarization angle. While for decoding stage, the detector will detect whether the bit is 0 or 1 after go through beam.

### D. Schematic Diagram

The basic polarization of light as one of the strategies to encode the photons is required by the laser that emitted numerous photons through a multi-stage process. The Stokes parameter, which is a value of a linear polarization angle, is used to determine the polarization state of light [15]. The Stokes vector is used to determine light polarization as it passes through an optical system. The following vector form can be used to represent the four Stokes parameters [16]:

$$s = \begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{bmatrix} \quad (2)$$

where  $s_0$  indicates total optical beam intensity,  $s_1$  indicates linear horizontal or vertical polarised light,  $s_2$  indicates +45° or -45° polarised light, and  $s_3$  indicates right circular or left circular polarised light.  $s_1$  and  $s_2$  will be affected by the simple rotation operator, but  $s_3$  will stay unaffected.

The polarizer determines the polarization states based on the intensity of the output beam. The intensity output can be determined using Malus's law, given as [17]–[19]:

$$I_o = I_i \cos^2(\theta) \quad (3)$$

where  $\theta$  is the secret polarization angle for the bits,  $I_o$  is the output intensity,  $I_i$  is the input intensity.

$$\frac{1}{2} \times [1 \cos(2\theta) \sin(\theta) 0] \times = \frac{1}{2} \times [1 + \cos(2\theta)] \quad (4)$$

where  $S$  is the input bit, Equation 2's condensed form is obtained as below:

$$\frac{1 + \cos(2\theta)}{2} = \cos^2 \theta \quad (5)$$

In this study, 90° of linearly polarized light represents bit 1 and 0° of linearly polarized light represents bit 0 in matrices, as shown in Table I.

TABLE I. STOKES VECTOR AND STATE OF POLARIZATION

Stokes Vector	State of polarization	Bit representation
$s: \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	Light that isn't polarized	-
$s: \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	Linearly polarized light at 0°	0
$s: \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$	Linearly polarized light at 90°	1

## IV. PRINCIPLE OF OPERATION WITH SCHEMATIC

Following is a detailed explanation of the full experimental process for the hardware over free space optics (FSO) illustrated in Fig. 3. The protocol's implementation was divided into three different phases which are encoding, rotation transformation and decoding phase.

### A. Encoding

Alice generates a state with a 0 linear polarization at the start of the protocol using a 0° polarizer. The polarization of light is accomplished by Alice and Bob using two sets of polarizers to change a bit into a quantum state referred as a qubit, described by a Mueller matrix [10].

$$M_{pol} = \frac{1}{2} \begin{bmatrix} 1 & \cos(2\theta) & \sin(2\theta) & 0 \\ \cos(2\theta) & \cos^2(2\theta) & \cos(2\theta) \sin(2\theta) & 0 \\ \sin(2\theta) & \cos(2\theta) \sin(2\theta) & \sin^2(2\theta) & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (6)$$

where  $\theta$  is the degree of polarisation angle that is set from 0° to 180° and  $M_{pol}$  is the polarizer's rotation.

The beam is directed towards the HWP at this point in order to modify the polarization of the beam using a secret polarization angle. The HWP is angled at a certain angle and is updated for each batch of bits.

### B. Rotation Transformations Stage

This section discusses the configuration of the unitary transformations used with half wave plates as well as the selection of the half wave plates' rotation angle ( $\theta$ ) with regard to the horizontal axis. This decision relies on the Mueller matrix formalism to guarantee that the half wave plate setup's input and output polarisation angles are equivalent.

1) *Half wave plate operation:* Fig. 2 shows the implementation of both polarizer and half wave plate. A half wave plate generates a 180° polarization shift between the fast and slow axes of a wave plate. A HWP that is installed on a rotator and will be rotating at a random angle decided upon by Alice and Bob is also included. Given that the Stokes vector has four dimensions, Mueller matrices, as shown by [20], can be used to characterize the optical components of the HWP device as four by four matrices. The HWP operation's rotation is shown as [21]:

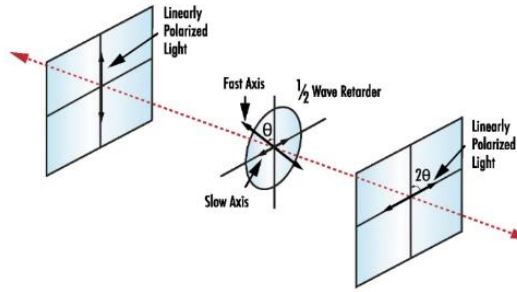


Fig. 2. Polarizer and half wave plate are employed [7]

$$M_{HWP} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(4\theta) & \sin(4\theta) & 0 \\ 0 & \sin(4\theta) & -\cos(4\theta) & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad (7)$$

Where

$$0^\circ < \theta < 180^\circ \quad (8)$$

### C. Decoding Stage

A beam combiner is used to combine the signals that come out of the channel. When the beam goes through the  $0^\circ$  and  $90^\circ$  polarizers, the detector will detect whether the bit is 0 or 1.

## V. HARDWARE IMPLEMENTATION

Table II lists the hardware components utilized in the simulation experiments, along with their descriptions. The schematic diagram is explained in the next section.

The hardware components listed in Table II are deployed into the simulation environment at this point. Depending on the requirements of the experiment, the hardware can be put in fiber optic or free space optic.

TABLE II. HARDWARE COMPONENT [15]

Components	Description
Laser	Light intensity detector
Beam Splitter	A 50/50 beam splitter is used to split a laser beam into equal-intensity beams.
Polarizer	Filter the laser beam to produce a $0^\circ$ or $90^\circ$ polarized light beam that indicates bit 0 or 1.
Half wave plate	Encrypt the laser's data and generate superposition states based on angles.
Light intensity detector	Check whether the photons are 0 or 1.

## VI. EXPERIMENTAL SETUP

Fig. 3 shows the implementation of HWP in message sharing procedure. Alice will have two half-wave plates (HWPs), whereas Bob will have one HWP. In our implementation, a linearly polarized laser serves as the photon source. Depending on whether the input bit is 0 or 1, polarizers are used to filter a beam of light to pass a  $0^\circ$  or  $90^\circ$  degree polarized beam [8]. Alice generates a state with a 0 linear polarization at the start of the protocol using a  $0^\circ$  polarizer.

The polarization technique generates a quantum bit (qubit) by encrypting the classical bits with photons. After encrypting the classical bits to the photons, Alice will use HWP-1 to execute a transformation at the angle  $\theta_A$ , this produces the superposition state,  $|\psi\rangle$ . The superposition states hold the secret message that Alice and Bob will communicate in quantum ways.

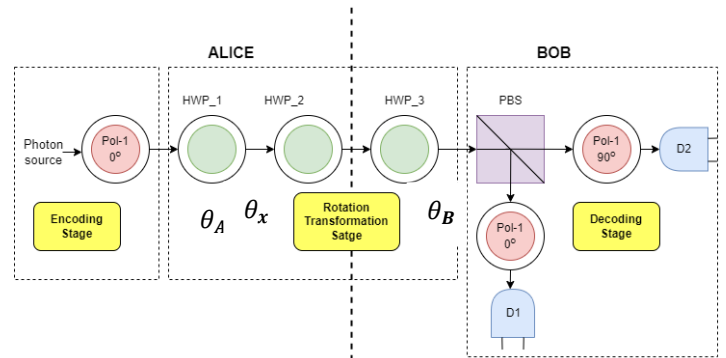


Fig. 3. Schematic of the experimental set-up over free space optics

In order to increase the security of the protocol, it is noted that HWP-1 at the angle  $\theta_A$  is the authentication key specified from  $0^\circ$  to  $45^\circ$ . The transformation coupled with the encoded bit is then applied by Alice using HWP-3, which will be set at an angle  $\theta_x = 0$  for bit 0 or  $\theta_x = 45$  for bit 1.

$$\text{Message } X \Rightarrow \text{bits} \in \{0,1\}$$

$$\{0,1\}^n \Rightarrow \begin{cases} |0\rangle, S_{out} = M_{HWP}(0^\circ) \times S_{in} \\ |1\rangle, S_{out} = M_{HWP}(45^\circ) \times S_{in} \end{cases} \quad (9)$$

where  $S_{in}$  is the input light's input state comes from the polarizer Stokes parameter sequence, which is represented as

$$S_{in} = \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad (10)$$

and  $S_{out}$  is the output light's input state from the polarizer Stokes parameter sequence:

$$S_{out} = \begin{bmatrix} S'_0 \\ S'_1 \\ S'_2 \\ S'_3 \end{bmatrix} \quad (11)$$

The input for this study is a  $0^\circ$  state. Consequently, the Stokes parameter's output is described by

$$S_{out} = M_{HWP}(\theta_x) \times S_{in} \quad (12)$$

After passing through the HWP, the light's polarization angle may be calculated [28,29] by:

$$\theta = \cos^{-1} \frac{S_{in}[1:3] \cdot S_{out}[1:3]}{\|S_{in}[1:3]\| \times \|S_{out}[1:3]\|} \quad (13)$$

where "." represents the multiplication operator and  $\|\cdot\|$  represents the norm operator.

Bob receives the optical beam from Alice that contains the message, and uses a HWP-3 set at an angle  $\theta_A$  to remove his transformation. After passing it via a polarization beam splitter,

Bob will receive in a beam polarizer at either  $0^0$  or  $90^0$  degrees. Accordingly, if  $\theta_x = 45$ , the output light will be vertically polarized and if  $\theta_x = 0$ , the output light will be horizontally polarized. Detectors will determine if the bit is 0 or 1.

### VII. EXPERIMENTAL SIMULATION

In this section, the proposed protocol's basic description is covered. First of all, the message that has been converted to binary code is encrypted into a quantum bit or also known as qubit. The proposed protocol is then used to communicate the qubit from Alice to Bob as follows.

1) *Let's say Alice gets a photon ready:* The first qubit,  $|0\rangle$  is what Alice wishes to transmit, and it is encrypted to the photons using the first defined angle from the first method,  $\theta_A = 30^0$ . Alice prepares photons according to Equation (10) and encrypts them by rotating HWP according to Equation (7):

$$\varphi_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(4(30)) & \sin(4(30)) & 0 \\ 0 & \sin(4(30)) & -\cos(4(30)) & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -0.5 \\ 0.8660254 \\ 0 \end{bmatrix} \quad (14)$$

2) Alice then performs the transformation corresponding with the encoded bit. The angle is set to  $\theta_x = 0^0$  if bit 0 is being transmitted or  $\theta_x = 45^0$  if bit 1 is being sent.

$$\varphi_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(4(0)) & \sin(4(0)) & 0 \\ 0 & \sin(4(0)) & -\cos(4(0)) & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -0.5 \\ -0.8660254 \\ 0 \end{bmatrix} \quad (15)$$

3) When Bob gets  $\varphi_2$ , he rotates it using the angle of the authentication key,  $\theta_A = -30^0$ , to decode it. The original message is delivered to Bob. The result of the  $\varphi_2$  represents the state  $|0\rangle$  that Alice sent.

$$\varphi_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(4(-30)) & \sin(4(-30)) & 0 \\ 0 & \sin(4(-30)) & -\cos(4(-30)) & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \times \begin{bmatrix} 1 \\ -0.5 \\ -0.8660254 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad (16)$$

### VIII. PSEUDO CODE

The algorithms in this work serve as an exact list of instructions that carry out the steps in the process of transmitting messages via optical devices. The steps include encoding, rotation of HWP and decoding stage. The pseudo-code of the protocol is detailed in Algorithm 1.

#### Algorithm 1

1. Notation:
2.  $R(\theta) \leftarrow$  is the rotation of HWP using Equation (6)
3.  $\theta_A \leftarrow$  Angle of HWP at Alice
4.  $\theta_x \leftarrow$  Angle of HWP at Alice either  $0^0$  or  $45^0$
5.  $\theta_B \leftarrow$  Angle of HWP at Bob
6. Alice prepares a photon using Equation (9)
7. Encoding Stage: Generate a photon to represent a qubit after the laser beam passed through a linear polarizer:
8. Alice generates a state with a  $0^0$  linear polarization using a  $0^0$  polarizer.
9. Rotation of HWP Stage
10.  $R(\theta_A) |\varphi_0\rangle = |\varphi_1\rangle$
11.  $R(\theta_x) |\varphi_1\rangle = |\varphi_2\rangle$
12. **if** bit 0 is being transmitted **then**
13.      $\theta_x = 0^0$
14. **else if** bit 1 is being sent **then**
15.      $\theta_x = 45^0$
16. **end if**
17.  $R(\theta_B) |\varphi_2\rangle = |\varphi_3\rangle$
18. Decoding Stage: The polarizer detects the photon's polarization states
19. Bob receives beam polarizer at either  $0^0$  or  $90^0$  degrees
20. **if**  $\theta_x = 0^0$  **then**
21.     the output light will be horizontally polarised
22. **else if**  $\theta_x = 45^0$  **then**
23.     the output light will be vertically polarised
24. **end if**
25. Detectors will determine if the bit is 0 or 1

### IX. CONCLUSION

Photonic polarization qubits are frequently employed in quantum computation and quantum communication due to their robustness in transmission and ease of manipulation. But the usage of optical devices such as half wave plate and polarizer might high costs. The more the number of half-wave plates implemented in an experiment, the more times needed to transmit the message. This is because of changing the polarization angle of an optical device for encoding purposes, more time is required. These circumstances have led to an increase in source redundancy, which in turn causes a rise in transmission time. After consideration, this experiment implements three half-wave plates only. Bear in mind that this experiment is to help to stimulate how was the quantum cryptography works on message sharing among the parties since it can describe quantum states in mathematical form. This study was written to explain that quantum experiments can be carried out using optical devices. And the stimulation shown can be used as a guideline for other researchers. This paper also contributes an alternative way through Python software to

simulate the function of each hardware component and uses the polarization of the quantum state to encrypt messages in a quantum communication environment.

#### ACKNOWLEDGMENT

This research was supported by Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS/1/2021/ICT11/UTHM/03/1). We also want to thank to the Government of Malaysia which provide MyBrain15 programme for sponsoring this work under the self-funded research grant and L00022 from Ministry of Science, Technology and Innovation (MOSTI).

#### REFERENCES

- [1] M. H. Adnan, Z. A. Zukarnain, and N. Z. Harun, "Quantum Key Distribution for 5G Networks: A Review, State of Art and Future Directions," *Future Internet*, vol. 14, no. 3. MDPI, Mar. 01, 2022. doi: 10.3390/fi14030073.
- [2] O. Lib and Y. Bromberg, "Quantum light in complex media and its applications," *Nat Phys*, vol. 18, pp. 986–993, Sep. 2022, doi: 10.1038/s41567-022-01692-y.
- [3] A. H. Dorrah and F. Capasso, "Tunable structured light with flat optics," *Science* (1979), vol. 376, no. 6591, p. eabi6860, 2022, doi: 10.1126/science.abi6860.
- [4] K. Prateek, F. Altaf, R. Amin, and S. Maity, "A Privacy Preserving Authentication Protocol Using Quantum Computing for V2I Authentication in Vehicular Ad Hoc Networks," *Security and Communication Networks*, vol. 2022, p. 4280617, 2022, doi: 10.1155/2022/4280617.
- [5] Y. Bloom, I. Fields, A. Maslennikov, and G. G. Rozenman, "Quantum Cryptography—A Simplified Undergraduate Experiment and Simulation," *Physics* (Switzerland), vol. 4, no. 1, pp. 104–123, Mar. 2022, doi: 10.3390/physics4010009.
- [6] Z. Shen and D. Huang, "A Review on Metasurface Beam Splitters," *Nanomanufacturing*, vol. 2, no. 4, pp. 194–228, Nov. 2022, doi: 10.3390/nanomanufacturing2040014.
- [7] L. Elite Optoelectronics Co., "The effect of placing a half wave plate and a polarizer in the optical path," Nov. 09, 2021. <http://m.s-laser.com/info/in-the-optical-path-what-is-the-effect-of-pla-63968296.html> (accessed Nov. 13, 2022).
- [8] Abushgra, Abdulbast A., and K. M. Elleithy, "A Shared Secret Key Initiated by EPR Authentication and Qubit Transmission Channels," *IEEE Access*, vol. 5, pp. 17753–17763, Aug. 2017, doi: 10.1109/ACCESS.2017.2741899.
- [9] N. Z. Harun, Z. A. Zukarnain, Z. M. Hanapi, and I. Ahmad, "Hybrid M-Ary in Braided Single Stage Approach for Multiphoton Quantum Secure Direct Communication Protocol," *IEEE Access*, vol. 7, pp. 22599–22612, 2019, doi: 10.1109/ACCESS.2019.2898426.
- [10] P. K. Verma, M. el Rifai, and K. W. C. Chan, "Multi-photon Quantum Secure Communication," 2019. [Online]. Available: <http://www.springer.com/series/4748>
- [11] S. Mandal et al., "Implementation of Secure Quantum Protocol using Multiple Photons for Communication," Aug. 2012, doi: 10.48550/arxiv.1208.6198.
- [12] M. S. Kang, Y. S. Kim, J. W. Choi, H. J. Yang, and S. W. Han, "Experimental quantum message authentication with single qubit unitary operation," *Applied Sciences* (Switzerland), vol. 11, no. 6, Mar. 2021, doi: 10.3390/app11062653.
- [13] B. Riggs et al., "Multi-Wavelength Quantum Key Distribution Emulation with Physical Unclonable Function," *Cryptography*, vol. 6, no. 3, Sep. 2022, doi: 10.3390/cryptography6030036.
- [14] M. Hayashi, "Finite-block-length analysis in classical and quantum information theory," *Proceedings of the Japan Academy Series B: Physical and Biological Sciences*, vol. 93, no. 3. Japan Academy, pp. 99–124, 2017. doi: 10.2183/pjab.93.007.
- [15] Darunkar and A. Bhagyashri, "Multi-photon Tolerant Quantum Key Distribution Protocol for Secured Global Communication," 2017.
- [16] X. Hu and C. R. Guzman, "Generation and characterization of complex vector modes with digital micromirror devices," Sep. 2021, doi: 10.1088/2040-8986/ac4671.
- [17] E. Hecht, "Optics Fifth Global Edition," Pearson, 2017. [https://www.academia.edu/44107964/OPTics\\_FiFTh\\_EdiTiON\\_GIObAl\\_EdiTiON](https://www.academia.edu/44107964/OPTics_FiFTh_EdiTiON_GIObAl_EdiTiON) (accessed Nov. 13, 2022).
- [18] Z. Li et al., "Three-Channel Metasurfaces for Multi-Wavelength Holography and Nanoprinting," *Nanomaterials*, vol. 13, no. 1, p. 183, Dec. 2022, doi: 10.3390/nano13010183.
- [19] E. Hecht, "Optics: A Contemporary Approach to Optics with Practical Applications and New Focused Pedagogy, Global edition," p. 725, 2017, Accessed: Nov. 13, 2022. [Online]. Available: <https://www.pearson.com/uk/educators/higher-education-educators/program/Hecht-Optics-Global-Edition-5th-Edition/PGM1095066.html>
- [20] J. J. Gil, R. Ossikovski, and I. S. José, "Physical Significance of the Determinant of a Mueller Matrix," *Photonics*, vol. 9, no. 4, Apr. 2022, doi: 10.3390/photonics9040246.
- [21] N. Z. Harun, Z. A. Zukarnain, Z. M. Hanapi, I. Ahmad, and M. F. Khodr, "Multiphoton quantum communication using multiple-beam concept in free space optical channel," *Symmetry* (Basel), vol. 13, no. 1, pp. 1–16, Jan. 2021, doi: 10.3390/sym13010066.

# Analysis of Medical Slide Images Processing using Depth Learning in Histopathological Studies of Cerebellar Cortex Tissue

Xiang-yu Zhang<sup>1</sup>, Xiao-wen Shi<sup>2</sup>, Xing-bo Zhang<sup>3\*</sup>

Artificial Intelligence, School of Information Science and Technology,  
Shandong Agriculture And Engineering University, JiNan, 250100, China<sup>1</sup>

Government Economic Management, School of Government,  
Beijing Normal University, Beijing, 100875, China<sup>2</sup>

Department of Internet of Things Engineering, School of Information Science and Technology,  
Shandong Agriculture and Engineering University,  
No. 866, Nongganyuan Road, Licheng District, Jinan District, JiNan, 250100, China<sup>3</sup>

**Abstract**—Today, with the advancement of science and technology, artificial intelligence evolves and grows along with human beings. Clinical specialists rely only on their knowledge and experience, as well as the results of complex and time-consuming clinical trials, despite the inevitable human errors of diagnosis work. Performing malignant and dangerous diseases, the use of machine learning makes it clear that the ability and capacity of these techniques are beneficial to help correctly diagnose diseases, reduce human error, improve diagnosis, and start treatment as soon as possible. In diseases, image processing and artificial intelligence is widely used in medicine and applied in stereological, histopathology. One of the essential activities for diagnosing the disease using artificial intelligence and machine learning is the fragmentation of images and classification of medical images, which is used to diagnose the disease with the help of images of the patient obtained from medical devices. In this article, we have worked on classifying medical histopathological images of brain tissue. The images are not of good quality due to sampling with standard equipment, and an attempt is made to improve the quality of the images by operating. Also, all images are segmented using the U-NET algorithm. In order to improve performance in classification, segmented images are used to classify images into two classes, normal and abnormal, instead of the images themselves. The images in the data set used in this study have a small number of images. Due to the use of a convolutional neural network algorithm to extract the feature and classify the images, more images are needed. Therefore, the data amplification technique to overcome this problem is used. Finally, the convolutional neural network has been used to extract features from images and classify fragmented images. Experimental results shown that the proposed method presented better performance compared to other existing methods.

**Keywords**—Image processing; fragmentation of images; machine learning; image classification; stereological; histopathology

## I. INTRODUCTION

Today, with advances in various fields such as engineering and medicine, organizations produce and collect large amounts of data every day. Hence, the need for data analysis to improve

an organization's processes, advance goals, speed things up, and reduce possible errors is felt. With the development of technology, there are significant advances in medicine that have increased data and information, so one of the topics in data analysis is to improve the quality-of-life process and community health, which is medical data. Machine learning can be used to improve performance and conclusions in medicine and reduce diagnostic errors. There are different types of medical data. Data can be text, image, video, and audio. In this article, the data used are medical images of the brain area taken from slides. In this way, brain cells are placed on a slide, one of the accessories of laboratory equipment, as an example for observation under a microscope. Then, microscopic observations are taken, and these images are used as data in this study.

One method of data analysis for medical images (brain area) is machine learning. Machine learning is the scientific study of algorithms and statistical models that computer systems use to perform a particular task without explicit instructions, patterns, and inference. This is known as a subset of artificial intelligence. Machine learning algorithms create a mathematical model based on sample data, known as "training data," to make predictions or decisions without explicitly planning for the task [1-4]. Using the built model, the machine learns that if new data enters the machine, the machine can use the built model to predict the new data and specify the desired label for each data. The purpose of this paper is to identify image tags using a machine [5-7].

The purpose of this study is to develop a method to perform malignant and dangerous diseases using machine learning techniques which the ability and capacity of these techniques are beneficial to help correctly diagnose diseases, reduce human error, improve diagnosis, and start treatment as soon as possible. In diseases, image processing and artificial intelligence is widely used in medicine and applied in stereological, histopathology. One of the essential activities for diagnosing the disease using artificial intelligence and machine learning is the fragmentation of images and classification of



medical images, which is used to diagnose the disease with the help of images of the patient obtained from medical devices.

Types of machine learning methods include supervised learning, semi-supervised learning, and unsupervised learning [8,9]. The data that will be used in this study are images of brain cells, all of which are labeled. So, learning that can be used for them has supervised learning. On the other hand, these data have two types of labels (normal and abnormal). As a result, the type of problem will be a binary classification problem. To learn, the machine must be able to extract features from photographs. An essential part of this study is the extraction of features from photographs and the fragmentation of photographs to examine microscopically the changes made by cerebellar cortex images in which the body changes [10,11]. The cellular neurons are comparable in different groups. The data used is raw and real, so that they will have some inconsistencies and problems. Most images are not of good quality and will confuse the machine. So, these problems must be fixed to improve performance and efficiency. On the other hand, it is necessary to prepare the data according to the conditions and the problem to the appropriate data to be injected into the desired algorithms. In general, it means that the data should be transformed in a way that the algorithm is acceptable and permissible for the implementation of algorithms, and the data should be cleared of any unnecessary errors and information to have the desired output. Data processing needs to be preprocessed. That will be the first step in this study. The most critical data preprocessing techniques include data cleansing, data integration, data reduction, and data transformation [12-15].

The next step is to segment the images. Fragmentation is a very related task in the analysis of medical images. Automatic separation of organs and structures of interest is often necessary to perform tasks such as visual augmentation, computer-aided detection, interventions, and the extraction of quantitative indicators from images [16-18]. Image segmentation, also called tagging, is the process of dividing an individual element of an image into a set of groups so that all the elements in a group have a common feature. In medicine, this common feature is usually that the elements belong to the same type of tissue or organ [19-23]. Different algorithms and methods can be used for segmentation. The purpose of segmentation is to simplify and change the representation of an image into something more understandable and easier to analyze. By segmenting the images, the borders and lines of the images will be determined, leading to better performance for the next step, which is to extract the features.

After preprocessing and segmentation, the feature extraction operation is performed on the data. Feature extraction is a process in which, by performing operations on data, it is obvious and determining features are determined. Feature extraction is done on the data used in this article to extract the features that are in the images. The machine is trained in these features, and based on the learning it has from these features, any other image except these images is provided to the machine so that the machine can correctly predict the label of the image without the label. There are different algorithms for feature extraction, which in this research will use a convolutional neural network algorithm. Deep learning

algorithms, especially convolutional neural networks, have quickly become a method for analyzing medical images. Deep network learning algorithms consist of many layers that input data while learning higher-level features. For example, it converts images to output (for example, the type of disease).

The most successful models for image analysis to date are convolutional neural networks [24-26]. In deep learning or hierarchy, several different layers are used to learn features. The more (deeper) this hierarchy of layers is, the more nonlinear properties are obtained. For this reason, this article has tried to use more layers. Of course, this statement has its limitations. After a certain number of layers, the system's performance does not improve, and new and more features are not trained. On the other hand, it is impossible to work directly on the input image to obtain or produce features containing more information. For this reason, deep learning with layers and hierarchies has been used.

Most dangerous diseases, such as cancer and the presence of tumors in the body's organs, are well diagnosed by examining and analyzing images obtained from sampled tissue samples. Therefore, it is possible to prevent the disease's progression and treat it as soon as possible by diagnosing it as quickly and accurately as possible. Therefore, a better and more accurate analysis of histological images can be beneficial to doctors and patients and can even be said to reduce the risk of death. Generally, histology specimens obtained in the operating room are processed by formalin and then embedded in paraffin to diagnose the disease. The tissue is then cut with a high-precision device and mounted on glass slides. The slides are stained with hematoxylin and eosin (H&E) to see the nucleus and cytoplasm. Finally, pathologists complete the diagnosis by visually examining histological slides under a microscope. However, in breast cancer diagnosis, histologists agree in 75% of cases and disagree in other cases [27-29]. Therefore, due to the complex nature of visible structures, traditional manual detection of tissue microstructure and general organization of nuclei in histological images is very time-consuming and can be subjective. As a result, computer-aided automation systems are essential to reduce specialists' workload by improving diagnostic efficiency and reducing the mentality of disease classification [30-33].

In this article, images of cerebellar tissue are used, the images used are of low quality, and the quality of the images should be normalized and then analyzed. In addition to the low quality of the images, when coloring the images, unfortunately, some color deposits are created in the image, which is quite similar to the cells in the texture and distorts the process. The color of the cells, on the other hand, is very similar to the background color of the tissue, slowing down the analysis. The challenges for segmentation and image classification are that the number of images is about 91, which is a minimal number and not enough to use deep learning to analyze images, and solutions had to be devised. On the other hand, an algorithm for segmentation should be selected that can perform well on histopathological images and has acceptable and good accuracy with low images. Regarding the classification algorithm, an algorithm that works well for histopathological medical images should be selected. It is expected that the quality of the images can be increased and the accuracy of the work can be

improved, and also that the images can be segmented with high accuracy. Most importantly, the results obtained by extracting features from the images using machine learning are better and more accurate than the physicians' manual method. The classification error can also be significantly reduced until new data or a histological image of the new patient can be presented. Correctly predict whether the new image has a normal or unnormal texture.

## II. PROPOSED ALGORITHM

Fig. 1 shows a flowchart of the general process and the proposed algorithm. In the first step, the image quality is improved, including two parts: color change and removing color deposits. Then, the images are divided into two training sets and experiments in the second step. In the third step, the images are segmented by the U-net algorithm. In the fourth step, due to the lack of images, data amplification technique training is applied to the images. In the fifth step, all images are pre-processed. Finally, feature extraction is performed on fragmented and amplified images using the convolution neural network algorithm and using these features. The convolution neural network classifies the images. In the following, each of these sections is examined.

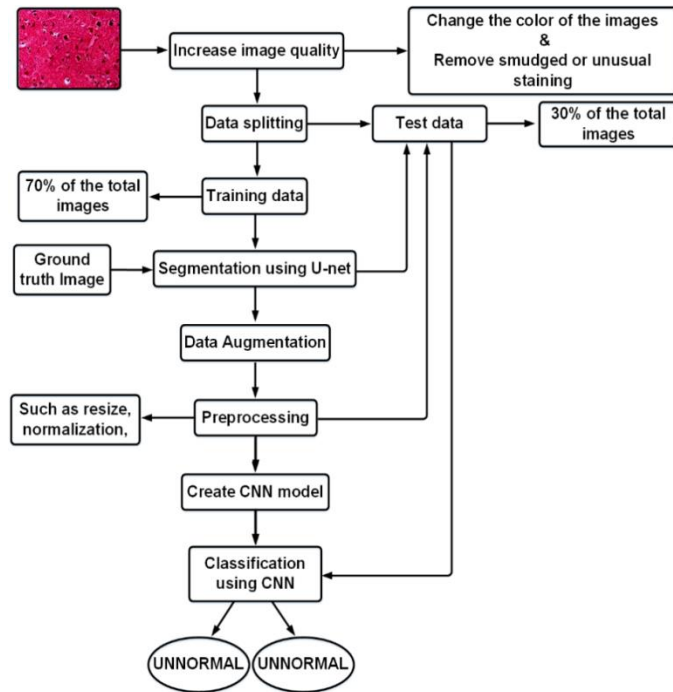


Fig. 1. The proposed algorithm

## III. PROPOSED PROCESS

In this section, the solution of the article is described in detail. According to the explanations in the previous chapters, the primary purpose of this article is to present a process for improvement in the classification of histopathological images of brain tissue using deep learning. Also, the following is a general introduction to the data set to understand the process better.

### A. Data Set

The data used in this article are brain tissue images. This dataset contains 91 histological images. All available images have labels, and there are two types of labels that are normal and abnormal. Each available image is only a member of one of these two tag classes. Class images do not overlap, meaning that an image can be a normal and non-normal class member and belongs to only one class. On the other hand, the number of images in these two classes is balanced, i.e., the number of images in the two classes is not much different from each other and is relatively evenly distributed in the two classes. For more familiarity with the available data, normal and unnormal image samples are shown in Fig. 2 and 3, respectively. As shown in the figures above, both images are very similar, and it will not be easy to distinguish between them. In histological images, in order to analyze the images, the cells in the histological images should be analyzed. These images have been used for the first time in this article, and no analysis has been done on these images before. Brain tissue is first sampled to image the data, then placed on a slide, a laboratory device, then stained by specialists to identify the cells better, and placed under a microscope and photographed. All the equipment used for sampling and photography is standard laboratory equipment, and no advanced and high-quality equipment is used because this sampling and imaging are done in a simple laboratory in the university. For this reason, the images are not of good quality.

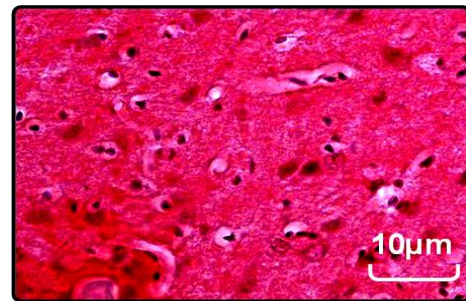


Fig. 2. Sample normal image

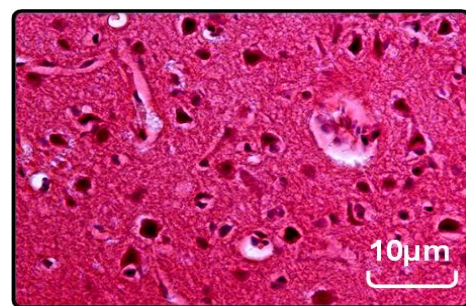


Fig. 3. Sample unnormal image

### B. Improve Image Quality

As shown in the previous images in Fig. 1 and 2, the background colors and the cells in the images are very similar. This may confuse the process, so to avoid this problem to increase the images' transparency and better distinguish the cells from the background, the color of the images has been changed using the Icy program. Fig. 4 shows an example of an image before and after a color change. The sediments in the

images are very similar to the actual cells. This causes many errors in the work process. For this purpose, the colored deposits in the images must be removed. To remove these sediments, Icy software and a paint program are used, and the

sediments are removed. Fig. 5 shows an example of images (a) before and (b) after sediment removal.

Due to a large number of images, all the steps performed in this step are applied to all images in the data set.

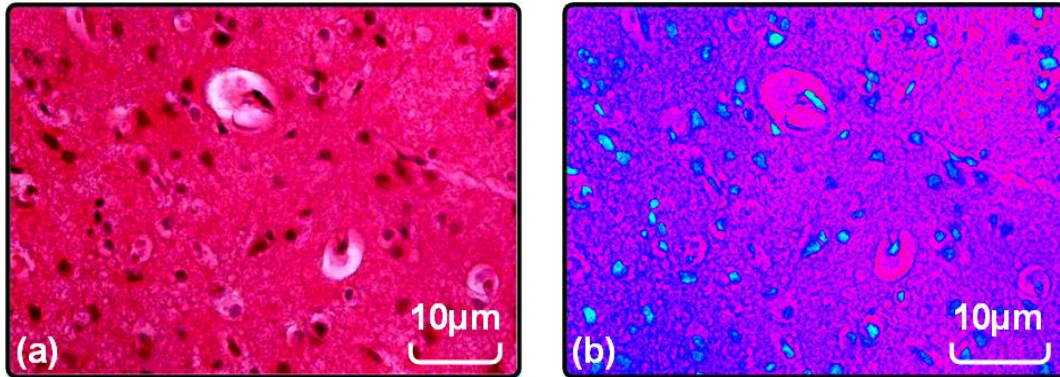


Fig. 4. Sample images, (a) before applying the color change, (b) after applying the color change

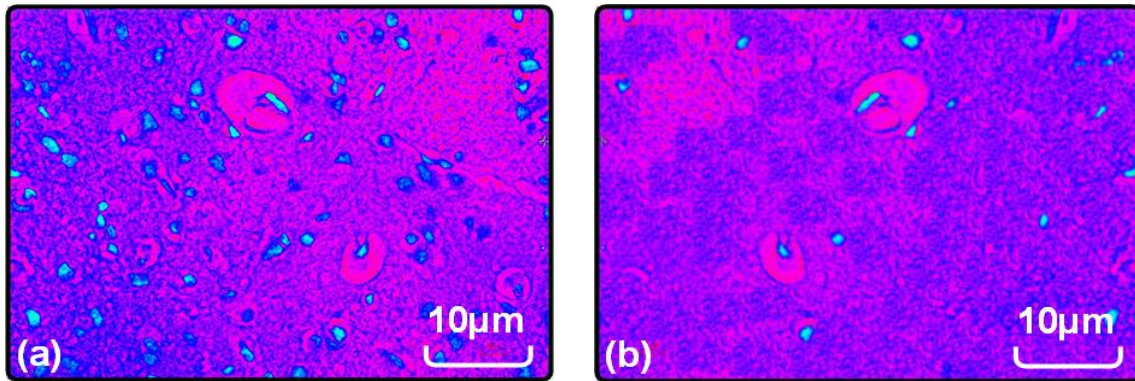


Fig. 5. Sample images, (a) before sediment removal, (b) after sediment removal

### C. Data Separation

After changing the color and removing color deposits, the images should be divided into two parts: training and testing. In all processes, the ratio of the number of training to the test is about 7 to 3, i.e., 70% of the normal data and 70% of the unnormal data are related to the training section, and the remaining 30% are related to the test section. The following is the number of data assigned to each class. The total images are equal to 91 images, of which 50 images belong to the normal class, and 41 images belong to the unnormal class. Out of a total of 91 images, 64 images are allocated to the training set and 27 images to the test set. Out of 64 images related to the training set, 34 are for normal class, and 30 are for unnormal class. Also, out of 27 images assigned to the test set, 16 are for normal class, and 11 are for unnormal class. As it turns out, there is a balance between the number of images in both classes, meaning that the distribution of images in the classes is relatively equal.

$$\text{Overall Image} \rightarrow 91 \begin{cases} \text{Normal} \rightarrow 50 \\ \text{Unnormal} \rightarrow 41 \end{cases}$$

$$\text{Training Data} \rightarrow 64 \begin{cases} \text{Normal} \rightarrow 34 \\ \text{Unnormal} \rightarrow 30 \end{cases}$$

$$\text{Teste Data} \rightarrow 27 \begin{cases} \text{Normal} \rightarrow 16 \\ \text{Unnormal} \rightarrow 11 \end{cases}$$

### D. U-Net Algorithm

This algorithm has a specific architecture, which is used in this article. The architecture is briefly described below. In 2015, Ronberberger et al. Proposed architecture for the segmentation of biomedical images [34]. This architecture is called U-net because of its U-shape. It is also called hourglass architecture. Convolution neural networks and deep learning techniques are not yet widely used because they have limitations due to the need for large amounts of data to teach complex models. However, the U-Net architecture does not require a large number of images due to the beneficial use of

data amplification techniques. Fig. 6 shows the architecture of this algorithm. This architecture has two contracting paths on the left and a Symmetric expanding path on the right. The contracting path is to extract the feature, and the expansion path is symmetrical to obtain the feature map. In general, this architecture has 23 layers of convolution. The contraction path follows the typical architecture of a convolutional network. This path involves repeated use of two 3×3 convulsions, each modified using a linear function (ReLU) and a maximum aggregation operation of 2×2 max using step 2 for downsampling. At each step of the descent, the number of feature channels doubles. Each step in the symmetric expansion path involves a mapping of the feature map, followed by 2×2 up sampling that halves the number of

channels and two 3×3 convulsions with the ReLU linear function for correction. Also, at each stage, the feature maps are concatenated in the opposite stage of the contraction path. Cutting at any stage is necessary due to the loss of boundary pixels in each convolution. In the last layer, a 1×1 convolution is used to map all 64 attribute vector components to the number of available classes. Data reinforcement is essential when the amount of data is small. Since there is often little data in medicine, the need for data reinforcement becomes sensitive in these cases. This is present in the U-Net architecture itself, and this has made U-Net work well for low-data datasets. All images are fragmented using this algorithm, and then the images generated by the output of the U-Net algorithm will be used.

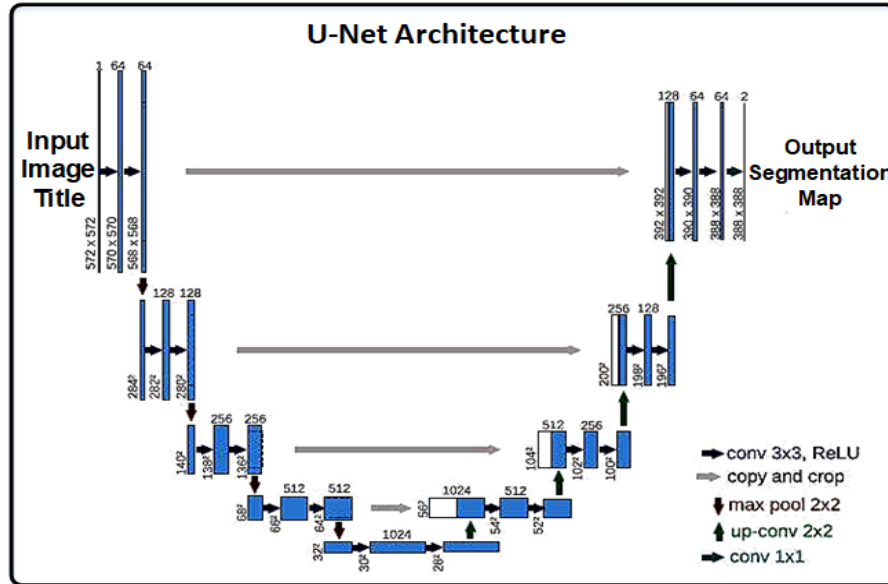


Fig. 6. U-Net architecture [34]

### E. Generate Mask Images

In order to segment images, there must be a tag for the images. The label in the segmentation is different from the label in the classification. In segmentation for system training, the system must be told which goal in the main image is to select and segment which part of the images. In image segmentation with the U-Net algorithm, the required tags are images called masks, which are given to the system to compare

the image and the mask corresponding to determine the purpose, segmentation, and selection. The real and mask sample images extracted by the excerpt are shown in Fig. 7.

As mentioned, these tags did not exist for this dataset. In this study, mask images are extracted using Icy software, a mask extraction program. This is done under the supervision of an expert so that the selection of essential areas would not be mistaken and the system could be properly trained.

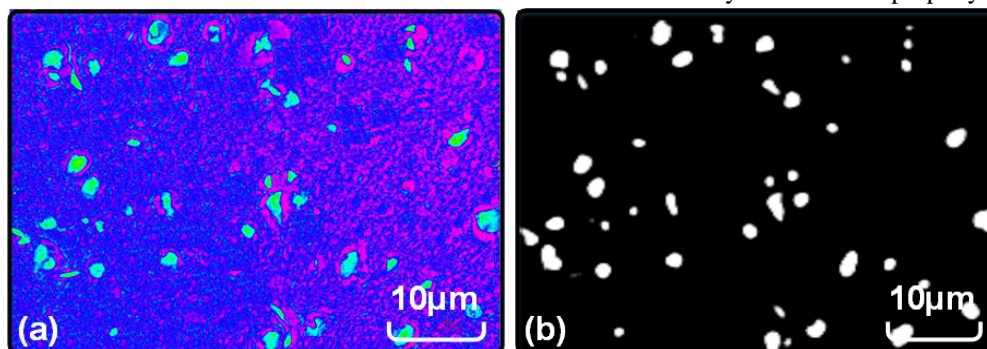


Fig. 7. Sample image (a) real, (b) mask

#### F. Data Augmentation

Improving the performance of deep neural networks is often directly related to the data available for training. In convulsive neural networks, if the amount of data is small, the system learning operation will not work well. More data is needed to improve the network. On the other hand, due to the lack of data in medical imaging, especially tissue imaging, due to complex sampling and imaging, and not much data collection is possible, the image enhancement technique will be beneficial. This technique allows the data to be amplified and the data set to be strengthened without requiring new images. This data will be based on the images in the current data set. By learning new properties from those images and the original images, the network gains the power to understand a broader range of each object and becomes so-called Generalized.

Therefore, it is clear that the selection of settings for the data amplification technique must be made carefully and follow the content of the data and the type of problem. Applicable trial and error enhance data enhancements for selected images. In this way, new images are generated by the

data amplification technique using different settings. Then the convolutional neural network is trained, and user settings in the data amplification technique are selected according to the improvement or non-improvement of the network. Modern deep-learning algorithms, such as convolutional neural networks, can learn features independent of their location in the image. However, amplifying and adding data by changing rotation, light, and color can further aid in learning independent features in the image. Eventually, the grid learns to recognize the object correctly if it is turned from left to right in the image or has little light and clarity. The critical point here is that data amplification should only be applied to training images, not test data, as this would compromise network performance. Fig. 8 shows the main figure. Fig. 9 also shows (a) the vertical flip and (b) the horizontal flip.

By applying the above changes to the entire tutorial images, about 50 images are added to each of the normal and unnormal image groups. The above changes do not pose a problem histologically because histologists can analyze images from any angle and direction.



Fig. 8. The main figure

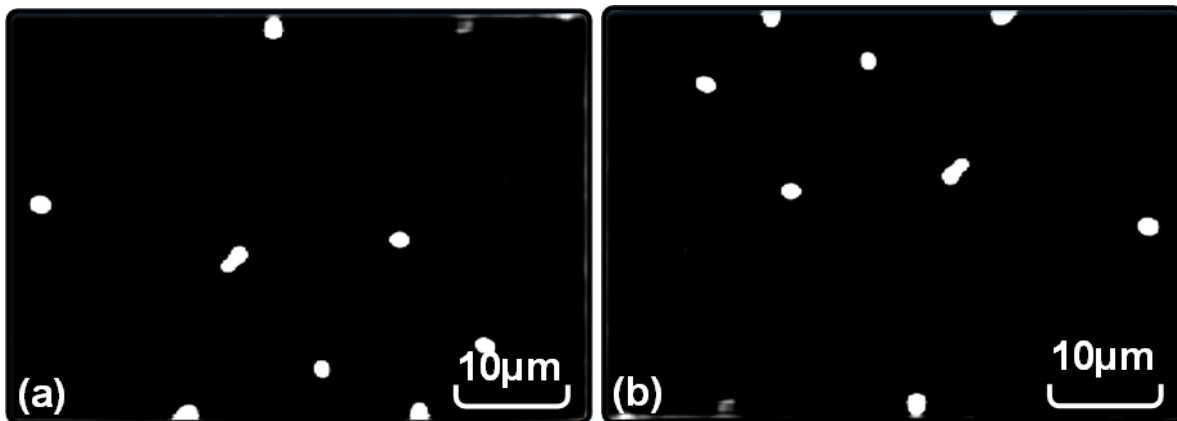


Fig. 9. Shows (a) the horizontal flip, (b) the vertical flip

#### IV. CONVOLUTION NEURAL NETWORK ARCHITECTURE

When a computer receives an image as input, it sees it as an array of numbers. The number of arrays depends on the size of the image (in pixels). For example, suppose a color image in JPG format with a size of 1280×960 pixels is given to the computer. In that case, its replacement array will have

3×1280×960 cells. The number 3 goes back to RGB and means that the image is in color. Each of these houses also has a number between 0 ~ 255. This number indicates the pixel intensity. The only available tools in the category of images using convolutional neural networks are such numbers. To understand and recognize complex images, such as the image

of a cell, the computer first recognizes the simpler features of that image, such as edges and curves.

The design of network layers can be different for each data, depending on the purpose pursued. The architecture designed in this article gave the best results for the data at hand. This architecture has 6 layers of convolution, and 5 layers of aggregation applied between the layers of convolution. In the last layer, a fully connected layer is used. The fully connected layer is responsible for calculating the score of categories (classes) to classify images. The filters used for the convolution layer have dimensions of  $5 \times 5$ , which include weights. The ReLU activator function is also used for the activator function. ReLU applies an activation function to each neuron, such as  $\max(0, x)$ , which sets the threshold at 0 (i.e., values  $x$  for negative values and 0 for negative values). This does not change the mass size from the previous step, i.e., the size of the output in this layer equals the output dimensions of the previous layer. This function does not change the dimensions of the output. The input layer is  $3 \times 224 \times 224$  images. The value of  $3 \times 224 \times 224$  are the dimensions of the images, and 3 also refers to the fact that the images are 3 channels, i.e., color images. Images enter the convolution neural network through the feed-forward stage. In the feed-forward stage, each filter is slid along the inlet's width and height, creating an Activation Map for that filter. As the filter slides across the inlet width, multiplying the point between the filter inputs by A stream is also performed. The result is passed through an activation function, and output is generated. In other words, at this stage, the input is multiplied by the weights, added to the bias, and entered into an activation function. This process continues until the end.

The initial weight of the grid is randomly weighted. Then Adam's reverse optimization algorithm is used to update the weights. The reverse algorithm is used to adjust the parameters based on their impact on the final result. This algorithm

calculates the error from the last layer. In order, layer by layer, the parameters are adjusted based on that error. That is, the weights and deviations are changed in such a way as to produce a result closer to the actual output with less error in the next iteration. CROSS ENTROPY is used as a loss function to determine the amount of error in the reverse algorithm, which aims to minimize the error. Optimizers and other activators can be used, each of which can be tried and tested. If each of these items is accurate, that item will be selected for use, but in most cases, they are not much different. In this study, it is concluded by trial and error to use the ReLU activator function and Adam optimizer. Fig. 10 shows the general structure of the architecture used.

As shown in the architecture shown in Figure 10, the first  $224 \times 224$  images enter the network. In the first layer, a convolution layer with a  $5 \times 5$  filter is applied to the image. The ReLU activator function is used, and the output is given to the aggregation layer with a  $2 \times 2$  filter and step 2, in which the dimensions of the images are halved. Then the output from the first aggregation layer is given to the second convolution layer again, and the same cycle continues. In designing convolutional neural network architecture, an important question is how many layers and what kind of sex is needed. Because the performance of architectures and layers is different for different data, like the other stages, trial and error have been used in this stage. Increasing the number of layers to a range improves neural network performance. After a range, increasing the number of layers reduces performance. This point can only be found by trial and error. The number of layers is also increased by trial and error. After this number of layers, as shown in Fig. 10, it is observed that the network performance does not improve. Even after the following few layers, it reduces the network performance. That is why the network is trained with 6 layers of convolution and ReLU, 5 layers of aggregation, and one fully connected layer.

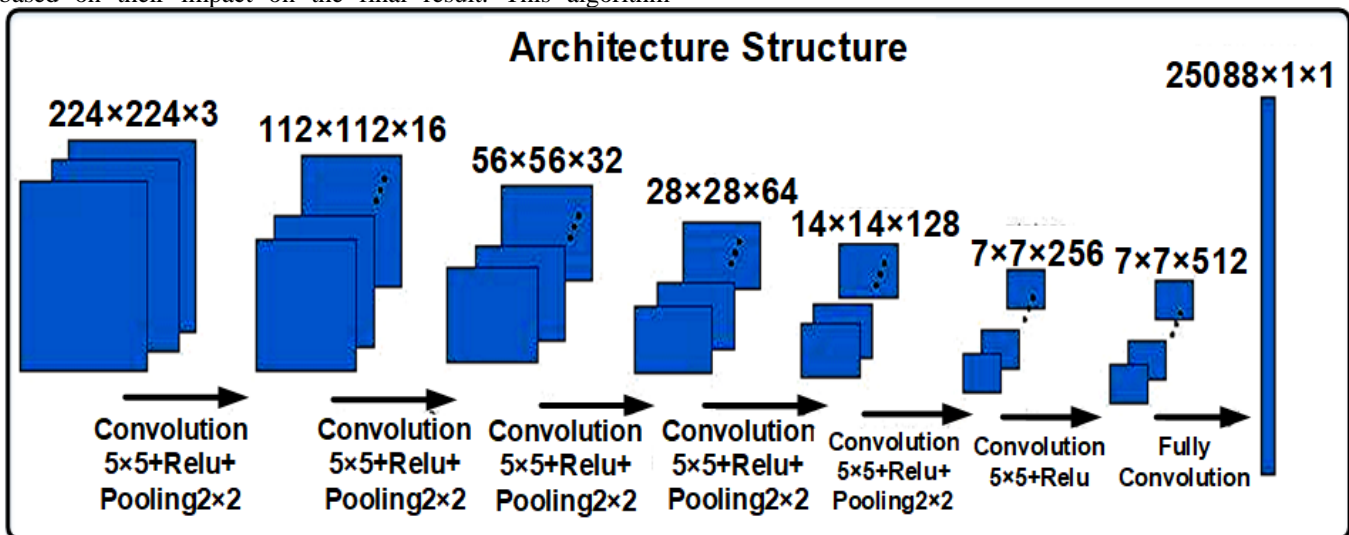


Fig. 10. General structure of architecture

## VI. RESULTS AND EVALUATION

In this section, evaluations and comparisons are provided to prove the efficiency of the proposed algorithm. For this purpose and to provide comprehensive results, the proposed method is reviewed in stages. In the following, the obtained results are discussed using evaluation criteria to measure the performance of the proposed process. The knowledge generated in the model learning stage should be analyzed in the evaluation stage to determine its value and then determine the efficiency of the model learning algorithm. These criteria can be calculated for the educational data set in the learning phase and the set of experimental records in the assessment phase. Evaluation of experimental data is often used because it shows the performance of the process. So, one of the most critical steps after designing and building a model or an algorithm is to evaluate its performance. Different evaluation criteria are used based on the existing algorithm and purpose. In this dissertation, two essential algorithms have been used, each with a separate evaluation. First, a brief description of the evaluation criteria used is given.

The results of the convolution neural network are described. In most previous works, the images entered the convolutional neural network directly after initial preprocessing. Alternatively, extracted through the neural network and then classified by another algorithm. The images are classified in several stages using a convolutional neural network, and the classification results are evaluated. All the results of these several stages are compared. The last step is related to the proposed process. The goal is to raise the standards of accuracy and reminder; the more these three criteria, the better the network performance. The criterion of accuracy obtained in all stages is in percentage.

On the other hand, the criteria of accuracy and numerical recall are between 0 and 1, which is closer to 1, i.e., the better the performance of that class. The data are randomly isolated six times for evaluation, and then all steps are performed. Finally, the results obtained from the repetition of six steps are averaged.

### A. Evaluation of Convolution Neural Network with Raw Images

In a reference study [24], the images entered the convolutional neural network for further analysis after applying techniques to reduce the number of images and select images. Because the data set, they used to have a lot of similar images. This study does not need to reduce the number of images because the data set used has few images. According to the reference, the images then entered the convolutional neural network [35]. First, the raw images entered the convolution neural network without any changes. Raw images are primary images without applying the segmentation algorithm and without applying the sediment removal and color change steps. The evaluation is performed 6 times on random data. Table I shows the accuracy, precision, and recall obtained in all these six steps. Also, in the end, the average of six steps will be considered.

In this table, the rows represent the evaluation steps, and the columns represent the evaluation criteria, which is the same

as the overall accuracy of the convolution neural network. Precision and Recall are also provided for each normal and abnormal class. Because both classes are equally important. Convulsive neural network in these conditions has an accuracy of about 76.14%. The normal class has a Precision of approximately 0.739 and a Recall of approximately 0.76. Also, the unnormal class has a Precision of approximately 0.76 and a Recall of approximately 0.73.

### B. Evaluation of Convolutional Neural Network with Fragmented Image

As mentioned before, in previous works, most images are classified using a convolutional neural network after a series of preprocessing to classify the images, especially histopathological images. In the solution proposed in this article, all images are segmented before the images are classified. Then, fragmented images are classified using a convolutional neural network. This step discusses the results obtained from the classification of fragmented images. The results for this process are according to Table II, in which also in this step, 6 random data are evaluated and finally averaged.

In this process, the accuracy is equal to 87.61%. Also, the precision and Recall of the normal class are equal to 0.871 and 0.91, respectively, and the precision and Recall of the unnormal class are equal to 0.915 and 0.846, respectively. As it is known, all the evaluation criteria in this process have improved compared to the previous process.

### C. Figures and Tables Evaluation of Convolution Neural Network with Segmented Image and Use of Data Amplification

In this series of evaluations, the quality of the images is first improved using the scale removal step and color change. The improved images are then segmented using the U-Net algorithm. Due to the lack of several images and in order to improve the performance of the convulsive neural network, the data amplification technique is used. The results obtained from the classification of this step of the images are according to Table III, also this step is repeated 6 times, and the average is taken.

The accuracy of this process is equal to 93.825%. Also, the precision and recall of the normal class are 0.912 and 0.978, respectively. On the other hand, the precision and recall of the unnormal class are equal to 0.979 and 0.9.

### D. Overall Evaluation Results

The following Table IV shows all the results of the evaluated steps. There is also a comparison chart in Fig. 11 to measure the accuracy of all steps.

Fig. 12 also provides a comparison chart for all steps and precision and recall criteria for each normal and abnormal class.

According to the diagram in Fig. 12 and Table IV, it can be seen that the accuracy obtained from the classification stage of fragmented images and the use of the data amplification technique has the highest percentage, approximately 93.82%. Also, according to Fig. 12 and Table IV, it is clear that the precision and recall of both normal and unnormal classes in the

classification of fragmented images and the use of data amplification are more than the other steps. Therefore, using the segmentation and data amplification technique in

histological medical images will have a good performance on the convolutional neural network.

TABLE I. RESULTS FROM THE CLASSIFICATION OF PRIMARY RAW IMAGES

-	Accuracy (%)	Precision (n)	Recall(n)	Precision(un)	Recall(un)
1	70	0.675	0.75	0.719	0.64
2	77.37	0.791	0.57	0.664	0.85
3	77.77	0.835	0.71	0.745	0.85
4	80.76	0.774	0.86	0.842	0.75
5	76.92	0.772	0.86	0.827	0.67
6	74.07	0.636	0.86	0.815	0.62
AVG	76.148	0.793	0.768	0.767	0.73
STDEV	3.69	0.074	0.116	0.07	0.102

TABLE II. RESULTS FROM IMPROVED AND SEGMENTED IMAGE CLASSIFICATION

-	Accuracy (%)	Precision (n)	Recall(n)	Precision(un)	Recall(un)
1	92.59	1	0.88	0.892	1
2	85	0.851	0.86	0.858	0.85
3	85.18	0.763	1	1	0.69
4	88.88	0.813	1	1	0.77
5	88.88	1	0.79	0.826	1
6	85.18	0.801	0.93	0.916	0.77
AVG	87.618	0.871	0.91	0.915	0.846
STDEV	3.05	0.103	0.82	0.72	0.129

TABLE III. RESULTS FROM THE CLASSIFICATION OF FRAGMENTED IMAGES AND THE USE OF DATA AMPLIFICATION TECHNIQUES

-	Accuracy (%)	Precision (n)	Recall(n)	Precision(un)	Recall(un)
1	96.3	1	0.94	0.943	1
2	92.59	0.869	1	1	0.85
3	92.59	0.869	1	1	0.85
4	92.59	0.869	1	1	0.85
5	92.59	1	0.93	0.934	1
6	92.59	0.869	1	1	0.85
AVG	93.825	0.912	0.978	0.979	0.9
STDEV	1.913	0.067	0.033	0.031	0.77

TABLE IV. OVERALL RESULTS OF ALL EVALUATION STEPS

-	Accuracy (%)	Precision (n)	Recall(n)	Precision(un)	Recall(un)
CNN	76.148	0.739	0.768	0.768	0.73
UNET+CNN	87.618	0.871	0.91	0.915	0.846
UNET+ Data augmentation+ CNN	93.825	0.912	0.978	0.979	0.9



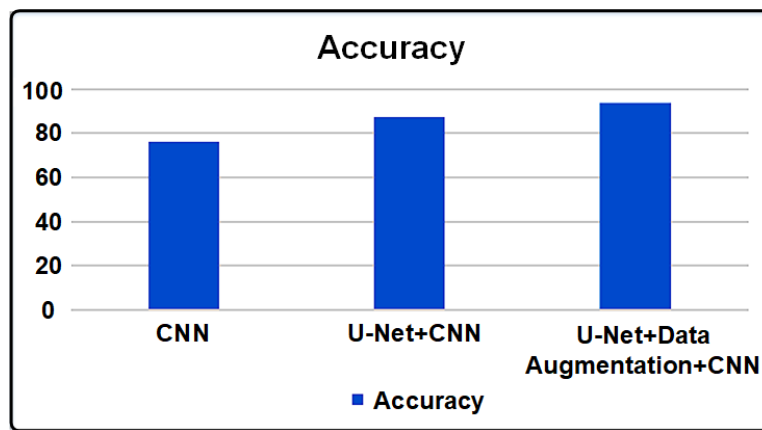


Fig. 11. Diagram related to the accuracy criterion of all evaluated steps

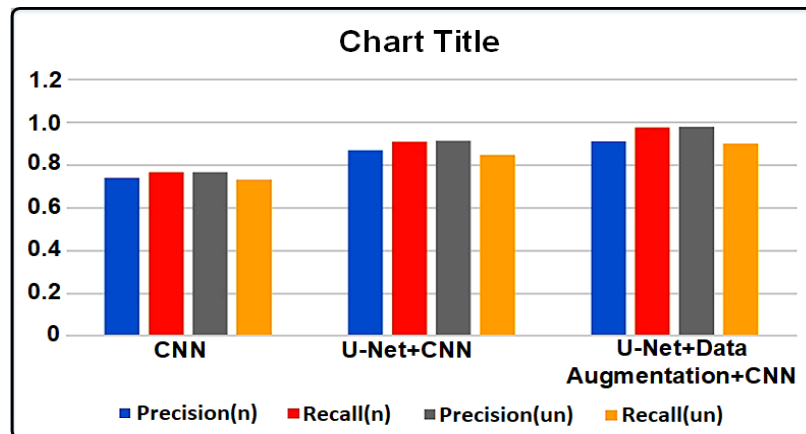


Fig. 12. Diagram of precision criteria and recall of all assessed steps

## VII. DISCUSSION

The accuracy, precision, and recall attained over the course of six steps are shown in this study's experimental results. To show the convolution neural network's overall accuracy, the evaluation process and evaluation standards are represented. Precision and Recall are also supplied for each normal and abnormal class. Because both classes are equally important. In these circumstances, the accuracy of the convulsive neural network is roughly 76.14%. Precision and Recall for the typical class are respectively 0.739 and 0.76. Also, the Abnormal class has a Precision of roughly 0.76 and a Recall of approximately 0.73.

According to the presented results, it can be seen that the accuracy obtained from the classification stage of fragmented images and the use of the data amplification technique has the highest percentage, approximately 93.82%. Moreover, the results of performance measurements indicates that the precision and recall of both normal and unnormal classes in the classification of fragmented images and the use of data amplification are more than the other steps. Therefore, using the segmentation and data amplification technique in histological medical images will have a good performance on the convolutional neural network.

## VIII. CONCLUSION

Today, deep learning has become one of the main techniques in the field of machine learning. They have solved many problems by using deep learning, and the use of deep learning has helped to improve the performance of many problems. For example, they use in-depth learning to diagnose various diseases, process medical images, social networking issues, etc. In this paper, deep learning is used to classify histological images. Images in different stages are classified using a neural network, as is clear from the results of the implementations. Improved image quality improved the performance of the U-Net segmentation algorithm. The U-Net algorithm also performed well for histopathological imaging.

On the other hand, the U-Net algorithm for small medical data sets can produce satisfactory results. The results obtained from the classification of fragmented images show that the segmentation of histological images improves the function of the convolutional neural network. The data amplification technique is also effective in this study because the number of images for training convolutional neural networks is minimal. Using the data amplification technique has led to better network learning. More images will ultimately help the process perform better. So, in general, it can be concluded that using segmented images and data amplification technique is helpful for histological images and causes better performance of the convolution neural network.

REFERENCES

- [1] S. Penchikala, Big data processing with apache spark, Lulu. com2018.
- [2] M. Juez-Gil, Á. Arnaiz-González, J.J. Rodríguez, C. López-Nozal, C. García-Osorio, Approx-SMOTE: Fast SMOTE for big data on apache spark, *Neurocomputing*, 464 (2021) 432-437.
- [3] V. Ankam, Big data analytics, Packt Publishing Ltd2016.
- [4] H. Kadkhodaei, A.M.E. Moghadam, M. Dehghan, Big data classification using heterogeneous ensemble classifiers in Apache Spark based on MapReduce paradigm, *Expert Systems with Applications*, 183 (2021) 115369.
- [5] A. Arooj, M.S. Farooq, A. Akram, R. Iqbal, A. Sharma, G. Dhiman, Big data processing and analysis in internet of vehicles: architecture, taxonomy, and open research challenges, *Archives of Computational Methods in Engineering*, (2021) 1-37.
- [6] J. Qiu, Q. Wu, G. Ding, Y. Xu, S. Feng, A survey of machine learning for big data processing, *EURASIP Journal on Advances in Signal Processing*, 2016 (2016) 1-16.
- [7] Q.-C. To, J. Soto, V. Markl, A survey of state management in big data processing systems, *The VLDB Journal*, 27 (2018) 847-872.
- [8] Y. Reddy, P. Viswanath, B.E. Reddy, Semi-supervised learning: A brief review, *Int. J. Eng. Technol*, 7 (2018) 81.
- [9] O. Chapelle, B. Scholkopf, A. Zien, Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews], *IEEE Transactions on Neural Networks*, 20 (2009) 542-542.
- [10] C. Brasko, K. Smith, C. Molnar, N. Farago, L. Hegedus, A. Balind, T. Balassa, A. Szkalitsy, F. Sukosd, K. Kocsis, Intelligent image-based in situ single-cell isolation, *Nature communications*, 9 (2018) 1-7.
- [11] A.J. Perez, M. Seyedhosseini, T.J. Deerinck, E.A. Bushong, S. Panda, T. Tasdizen, M.H. Ellisman, A workflow for the automatic segmentation of organelles in electron microscopy image stacks, *Frontiers in neuroanatomy*, 8 (2014) 126.
- [12] W. Sun, Z. Cai, Y. Li, F. Liu, S. Fang, G. Wang, Data processing and text mining technologies on electronic medical records: a review, *Journal of healthcare engineering*, 2018 (2018).
- [13] Z. Guan, T. Ji, X. Qian, Y. Ma, X. Hong, A survey on big data preprocessing, 2017 5th Intl Conf on Applied Computing and Information Technology/4th Intl Conf on Computational Science/Intelligence and Applied Informatics/2nd Intl Conf on Big Data, Cloud Computing, Data Science (ACIT-CSII-BCD), *IEEE*, 2017, pp. 241-247.
- [14] M.Z. Al-Taie, S. Kadry, J.P. Lucas, Online data preprocessing: a case study approach, *International Journal of Electrical and Computer Engineering*, 9 (2019) 2620.
- [15] S. Singhal, M. Jena, A study on WEKA tool for data preprocessing, classification and clustering, *International Journal of Innovative technology and exploring engineering (IJTe)*, 2 (2013) 250-253.
- [16] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, 2016 fourth international conference on 3D vision (3DV), *IEEE*, 2016, pp. 565-571.
- [17] Z. Zhu, C. Liu, D. Yang, A. Yuille, D. Xu, V-NAS: Neural architecture search for volumetric medical image segmentation, 2019 International conference on 3d vision (3DV), *IEEE*, 2019, pp. 240-248.
- [18] H.R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayashi, M. Oda, M. Fujiwara, K. Misawa, K. Mori, An application of cascaded 3D fully convolutional networks for medical image segmentation, *Computerized Medical Imaging and Graphics*, 66 (2018) 90-99.
- [19] E. Smistad, T.L. Falch, M. Bozorgi, A.C. Elster, F. Lindseth, Medical image segmentation on GPUs—A comprehensive review, *Medical image analysis*, 20 (2015) 1-18.
- [20] A. Eklund, P. Dufort, D. Forsberg, S.M. LaConte, Medical image processing on the GPU—Past, present and future, *Medical image analysis*, 17 (2013) 1073-1094.
- [21] T. Kalaiselvi, P. Sriramakrishnan, K. Somasundaram, Survey of using GPU CUDA programming model in medical image analysis, *Informatics in Medicine Unlocked*, 9 (2017) 133-144.
- [22] I.R.I. Haque, J. Neubert, Deep learning approaches to biomedical image segmentation, *Informatics in Medicine Unlocked*, 18 (2020) 100297.
- [23] H. Seo, M. Badii Khuzani, V. Vasudevan, C. Huang, H. Ren, R. Xiao, X. Jia, L. Xing, Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications, *Medical physics*, 47 (2020) e148-e167.
- [24] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, *Medical image analysis*, 42 (2017) 60-88.
- [25] M.I. Razzak, S. Naz, A. Zaib, Deep learning for medical image processing: Overview, challenges and the future, *Classification in BioApps*, (2018) 323-350.
- [26] J. Ker, L. Wang, J. Rao, T. Lim, Deep learning applications in medical image analysis, *Ieee Access*, 6 (2017) 9375-9389.
- [27] Y.S. Vang, Z. Chen, X. Xie, Deep learning framework for multi-class breast cancer histology image classification, *International conference image analysis and recognition*, Springer, 2018, pp. 914-922.
- [28] Y. Guo, H. Dong, F. Song, C. Zhu, J. Liu, Breast cancer histology image classification based on deep neural networks, *International conference image analysis and recognition*, Springer, 2018, pp. 827-836.
- [29] M.B.H. Thuy, V.T. Hoang, Fusing of deep learning, transfer learning and gan for breast cancer histopathological image classification, *International Conference on Computer Science, Applied Mathematics and Applications*, Springer, 2019, pp. 255-266.
- [30] S. Vesal, N. Ravikumar, A. Davari, S. Ellmann, A. Maier, Classification of breast cancer histology images using transfer learning, *International conference image analysis and recognition*, Springer, 2018, pp. 812-819.
- [31] S. Khan, N. Islam, Z. Jan, I.U. Din, J.J.C. Rodrigues, A novel deep learning based framework for the detection and classification of breast cancer using transfer learning, *Pattern Recognition Letters*, 125 (2019) 1-6.
- [32] C. Munien, S. Viriri, Classification of hematoxylin and eosin-stained breast cancer histology microscopy images using transfer learning with EfficientNets, *Computational Intelligence and Neuroscience*, 2021 (2021).
- [33] H. Cao, S. Bernard, L. Heutte, R. Sabourin, Improve the performance of transfer learning without fine-tuning using dissimilarity-based multi-view learning for breast cancer histology images, *International conference image analysis and recognition*, Springer, 2018, pp. 779-787.
- [34] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234-241.
- [35] M. Lorentzon, Feature extraction for image selection using machine learning, 2017.

# The Cloud-powered Hybrid Learning Process to Enhance Digital Natives' Analytical Reading Skills

Sakolwan Napaporn<sup>1</sup>, Sorakrich Maneewan<sup>2</sup>, Kuntida Thamwipat<sup>3</sup>, Vitsanu Nittayathammakul<sup>4\*</sup>

Division of Learning Innovation and Technology-Faculty of Industrial Education and Technology,  
King Mongkut's University of Technology Thonburi, Bangkok, Thailand<sup>1</sup>

Department of Educational Communications and Technology-Faculty of Industrial Education and Technology,  
King Mongkut's University of Technology Thonburi, Bangkok, Thailand<sup>2,3</sup>

Program in Learning Innovation-Faculty of Industrial Education,  
Rajamangala University of Technology Suvarnabhumi, Suphanburi, Thailand<sup>4</sup>

**Abstract**—Analytical reading is a necessary cognitive skill for advancing to other skills required in the digital age. Thailand is focused on the instructional development and use of digital media to enhance the digital natives' analytical reading skills, which will assist learners of all ages in effectively and quickly adapting to changes in the digital environment. After the COVID-19 pandemic situation, educational institutions in Thailand have begun to embrace a hybrid learning approach like never before. The limitations of the existing learning process for boosting digital natives' analytical reading skills are the lack of integration between reading techniques, hybrid pedagogies, and emerging learning technologies to enhance learners' seamless learning experiences. Thus, this study aims to propose the Cloud-powered Hybrid Learning process (Cp-HL process) to enhance digital natives' analytical reading skills. This study consisted of two main stages in the research methodology: 1) learning process development; and 2) learning process evaluation. The developed Cp-HL process had four main learning phases: (1) preparation for hybrid learning; (2) presentation for interactive learning; (3) practice with analytical reading; and (4) progress reports on analytical reading skills. All the experts agreed that the newly developed Cp-HL process performed extremely well in terms of overall suitability.

**Keywords**—Hybrid learning; cloud-powered learning tools; learning process; analytical reading skills; digital natives

## I. INTRODUCTION

Thailand's Key National Strategies for Human Capital Development and Strengthening focused on innovative-based instructional development and the use of digital media to help the Thai people become more moral, skilled, disciplined, kind, analytical readers, and able to "know, get, and adapt" to new technology, so that no one was left behind [1].

Analytical reading skills are high-level cognitive skills that allow a person to break down an intellectual or conceptual whole into parts for their individual study. This helps a person think, seek information, analyze information, solve problems based on that information, and make decisions in everyday life with more clarity [2–4]. The cooperative integrated reading and composition (CIRC) technique based on cooperative learning, which is intended to develop reading, writing, and other language skills, is one way to develop analytical reading abilities [5–7].

Digital natives become acquainted with digital media, digital technologies, and cloud computing at a youthful age and regard digital devices as an integral and necessary part of their lives. Many teenagers and children in developed countries are considered "digital natives" because they communicate and learn primarily through computers, social networking sites, and instant messaging. Thus, digital natives think, learn, and understand the world differently than people who have not been exposed to digital technologies [8–10].

Previous studies found that Hybrid Learning (HL) strategies are more successful than conventional learning strategies [11–13]. The success of implementing the HL strategy (experimental class) demonstrates that the hybrid strategy is appropriate for millennial students (generation Z as digital natives), or students born between 1995 and 2010. This supports several previous studies [14–15]. The HL strategy is superior to the typical classroom or face-to-face class in terms of active learning, student learning results, and learning satisfaction [13–15].

The limitations of the existing learning model for boosting digital natives' analytical reading skills are the lack of integration between characteristics of digital learners, digital reading techniques, hybrid pedagogies, and emerging learning technologies in a hybrid learning mode to enhance learners' seamless learning experiences [2, 16].

Thus, this study aims to propose the Cloud-powered Hybrid Learning process (Cp-HL process) to enhance digital natives' analytical reading skills. This article presents the research objectives, literature review on hybrid learning, cloud-powered learning tools, the CIRC technique, digital natives' analytical reading skills, the developed conceptual framework, research methodology, key findings, conclusions, and discussions.

## II. RESEARCH OBJECTIVES

- 1) To develop the Cp-HL process in order to enhance digital natives' analytical reading skills.
- 2) To evaluate the suitability of the Cp-HL process to enhance digital natives' analytical reading skills.

\*Corresponding Author.

### III. LITERATURE REVIEW

#### A. Hybrid Learning (HL)

After the COVID-19 pandemic, the one new normal in instructional design is Hybrid Learning (HL) [13, 17–18, 32]. HL, also known as blended learning or flipped learning, may vary widely in instructional design, execution, and learning support. "Face-to-face (F2F)" instruction includes those courses in which 0 to 29% of the content is delivered online, which includes both traditional and web-facilitated courses [13, 19–20]. The remaining alternative online courses are defined as having at least 80% of the course content delivered online. Whereas HL is defined as having 30 to 79% of the course content delivered online via F2F [13, 20]. This pedagogy combines the best components of both traditional classroom instruction and online classroom instruction, creating a flexible and dynamic learning environment. In a hybrid learning setting, students have the opportunity to attend classes in person, participate in online learning activities, and access digital resources and tools [19–20, 32–33].

#### B. Cloud-powered Learning Tools

Cloud-powered learning tools enable instructors to set up hybrid learning spaces where students can communicate and work together from anywhere using their digital devices, both in online learning mode and in on-site learning mode. These tools can be divided into four categories [13, 17–18, 21–29]: 1) social communication tools, such as Facebook®, LINE®, Telegram®, and Discord®; 2) browsing and seeking tools, such as Google Chrome®, Firefox®, Microsoft Edge®, and Safari®; 3) interactive presentation tools, such as Zoom®, Google Meet®, Microsoft Teams®, and ClassPoint®; 4) analytics and evaluative tools, such as ClassDojo®, Kahoot!®, Google Forms®, and Google Analytics®

#### C. CIRC Technique

The cooperative integrated reading and composition (CIRC) technique is the most effective and relevant technique for teaching students to read. CIRC derives from cooperative learning, which facilitates students' comprehension of the assigned reading material. Students are instructed to work in groups to achieve the objective of reading descriptive text and photos [5–7, 30].

#### D. Digital Natives' Analytical Reading Skills

Digital natives' analytical reading skills are rooted in the principles of analytical reading described in Dechant (1982) [31]. Previous studies about the level of understanding learners have of texts (reading) can be summarized into three components discussed as follows: 1) Comprehension reading is the level at which the reader can clearly explain, remember, and recall the idea or the details of the readable information. 2) Interpretative reading is a level of insight that is deeper than understanding the meaning of a word. Readers must be able to interpret or understand what the author does not directly address. 3) Analytical reading can reveal the latent meaning, understanding, aims, opinions, attitudes, and language level that the author has available. When they summarized what they read, they said what they liked and didn't like about it. When they retold a story, they used the information from their writing to make a conceptual schema.

Thus, digital natives' analytical reading should start with reaching and comprehending the elements of the media symbol systems, especially digital media, and then considering their validity and analyzing the message of the story in digital media by checking the elements to see how those elements make sense in terms of being intelligible, justifiable, or practicable or not. Digital natives' analytical reading essentially disassembles the entire intrinsic elements of the story under specific media symbol systems and examines how the elements relate to one another to respond to the story via interacting digital media symbol systems [2–4, 8–10]. Elements of analytical reading for digital natives can be put into five main categories [2–10]: 1) the media symbol system; 2) content; 3) language; 4) purpose; and 5) structure.

### IV. CONCEPTUAL FRAMEWORK

As shown in Fig. 1, the conceptual framework of this research used three ideas to come up with the Cp-HL process to help digital natives become better at analytical reading.



Fig. 1. Conceptual framework

Fig. 1 shows the main concepts and variables of the study to enhance the analytical reading skills of digital natives. The conceptual framework is made up of three parts:

(1) The basics of novel concept formation were divided into three concepts: *a*) Cloud-powered learning tools, which have four types: *i*) social communication tools; *ii*) browsing and seeking tools; *iii*) interactive presentation tools; *iv*) analytics and evaluative tools. *b*) Hybrid learning, which has two main learning modes: *i*) online learning mode; and *ii*) on-site learning mode. *c*) The CIRC technique, which has four main learning steps: *i*) instructor presentation; *ii*) team practice; *iii*) reading testing; *iv*) team recognition.

(2) The manipulated variable was the Cp-HL process to enhance digital natives' analytical reading skills into four main learning phases: *a*) preparation for hybrid learning (online learning mode); *b*) presentation for interactive learning (online learning mode); *c*) practice with analytical reading (on-site learning mode); *d*) progress reports in analytical reading skills (on-site learning mode)

(3) The dependent variable was digital natives' analytical reading skills, which had three components: *a*) identifying the intrinsic elements of the story and media symbol systems; *b*) analyzing the message of the story and media symbol systems; *c*) responding to the story and media symbol systems.

### V. RESEARCH METHODOLOGY

#### A. Phase I: Development of the Cp-HL Process in Order to Enhance Analytical Reading Skills for Digital Natives

In the first phase, we conducted research in the following steps: 1) a scoping review on the development of the Cp-HL process to enhance digital natives' analytical reading skills (i.e., hybrid learning, cloud-powered learning tools, CIRC

technique, analytical reading skills, and digital natives) in Thai Journals Online (ThaiJO), Education Resources Information Center (ERIC), and Scopus online databases published between 2017 and 2022 using content analysis of the research article's texts; 2) synthesis and construction of the conceptual framework; and 3) development of the Cp-HL process to enhance digital natives' analytical reading skills according to the Sloan Consortium's HL guidelines [20], which specifies a proportion of program content delivered online ranging from 30 to 79% as the most widely accepted and popular in Thailand.

### B. Phase II: Evaluation of the Suitability of the Cp-HL Process to Enhance Analytical Reading Skills for Digital Natives

In the second phase, we conducted research in the following steps: 1) creating research tools for evaluating the

suitability of the Cp-HL process; 2) proposing the Cp-HL process to nine experts in educational technology, Thai language, and instructional system design (three experts per side), each of whom holds a PhD or its equivalent and has at least three years of relevant experience; and 3) analyzing the results of the validation and evaluation of the Cp-HL process by the mean and standard deviation (S.D.). The suitability evaluation questionnaire was built on a five-point Likert scale (5 = strongly agree; 4 = agree; 3 = neither agree nor disagree; 2 = disagree; 1 = strongly disagree).

## VI. RESEARCH FINDINGS

### A. The Developed Cp-HL Process to Enhance Digital Natives' Analytical Reading Skills

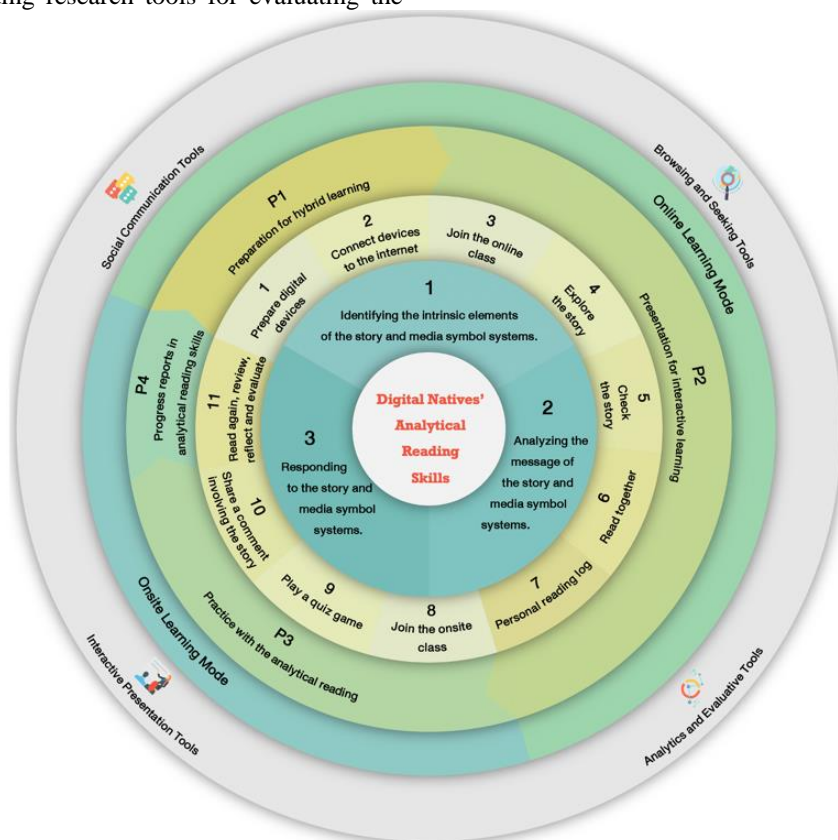


Fig. 2. The Cloud-powered Hybrid Learning process (Cp-HL process) to enhance digital natives' analytical reading skills

The developed Cp-HL process to enhance digital natives' analytical reading skills is illustrated in Fig. 2. This learning process consists of four main processes (4Ps) and 11 learning steps, as follows:

1) *P1: Preparation for hybrid learning*: In the first phase of the learning process, learners play an important role in self-directed learning; this phase has two learning stages under the online learning mode: a) Prepare digital devices (e.g., desktop, laptop, tablet, smartphone, and wearables); b) Connect devices to the Internet over a 3G, 4G, 5G, or Wi-Fi connection.

Thus, at this stage, the role of the instructor in hybrid learning management, as follows: a) Set up learning objectives, learning experiences (e.g., learning contents, pedagogies, and learning technologies), and learning assessments (e.g., formative assessment, summative assessment) that directly address the desired learning outcome, which is digital natives' analytical reading skills; and b) Encourage informal and formal communications using a variety of social communication tools (e.g., Facebook®, LINE®, Telegram®, and Discord®) and keep their frequency with learner groups in preparation for hybrid learning so instructors can open the learning cycles.

2) *P2: Presentation for interactive learning:* In the second phase of the learning process, learners play an important role in active learning; this phase has five learning stages under the online learning mode: *a)* Join the online class via browsing and seeking tools (e.g., Google Chrome®, Firefox®, Microsoft Edge®, and Safari); *b)* Explore the story on the interface display of interactive presentation tools (e.g., Zoom®, Google Meet®, Microsoft Teams®, and ClassPoint®); *c)* Check the story to identify the story's intrinsic elements and media symbol systems; *d)* Read together to analyze the story's message and media symbol systems; *e)* Personal reading log via browsing and seeking tools.

Thus, at this stage, the role of the instructor in an interactive presentation, as follows: *a)* Make it clear that instructors have high expectations for learning outcomes, that there are requirements for learning tasks and activities, and that there are rules for how to talk and interact in the classroom, both in online and on-site learning modes; and *b)* Respect diverse student abilities and learning styles to create a safe and engaging hybrid learning environment that allows for multiple responses. This will help make the classes more engaging (cognitively, socially, and emotionally) for both online and on-site learning.

3) *P3: Practice with the analytical reading:* In the third phase of the learning process, learners play an important role in collaborative learning; this phase has three learning stages under the on-site learning mode: *a)* Join the onsite class via analytics and evaluative tools (e.g., ClassDojo®, Kahoot!®, Google Forms®, and Google Analytics®); *b)* Play a quiz game via analytics and evaluative tools; *c)* Share a comment involving the story using the Think-Pair-Share (TPS) technique.

Thus, at this stage, the role of the instructor in practice with the analytical reading is very important, as follows: *a)* Respect diverse student opinions and thinking styles to create a safe and engaging hybrid learning environment that allows for multiple responses; *b)* Employ student peer assessment to promote collaborative learning skills; *c)* Develop reciprocity and cooperation among students to enhance peer support and collaboration.

4) *P4: Progress reports in analytical reading skills:* In the fourth phase of the learning process, learners play an important role in reflective learning under the on-site learning mode. This step emphasizes that the students read again, review, reflect, and evaluate the digital media for identifying the intrinsic elements of the story and media symbol systems, analyzing the message of the story and media symbol systems, and responding to the story and media symbol systems. Thus, at this stage, the role of the instructor in the progress reports of analytical reading skills, as follows: *a)* Employ student self-assessment to promote self-regulated learning skills, such as time management and learning progress monitoring; *b)* Give them quick feedback, and encourage them when they need it; *c)* Always end each lesson with a summary of what was learned and a look at what will be taught in the next lesson.

## B. The Suitability of the Cp-HL Process to Enhance Digital Natives' Analytical Reading Skills

TABLE I. THE EVALUATION OF THE SUITABILITY OF THE CP-HL PROCESS TO ENHANCE DIGITAL NATIVES' ANALYTICAL READING SKILLS

Items of evaluation	Result		Suitability Level
	Mean	S.D.	
<b>1. P1: Preparation for hybrid learning</b>			
1.1 Prepare digital devices	5.00	0.00	Strongly agree
1.2 Connect devices to the Internet	4.66	0.50	Strongly agree
<b>2. P2: Presentation for interactive learning</b>			
2.1 Join the online class	5.00	0.00	Strongly agree
2.2 Explore the story	5.00	0.00	Strongly agree
2.3 Check the story	4.66	0.50	Strongly agree
2.4 Read together	5.00	0.00	Strongly agree
2.5 Personal reading log	4.44	0.73	Agree
<b>3. P3: Practice with the analytical reading</b>			
3.1 Join the onsite class	5.00	0.00	Strongly agree
3.2 Play a quiz game	4.44	0.73	Agree
3.3 Share a comment involving the story	4.66	0.50	Strongly agree
<b>4. P4: Progress reports in analytical reading skills</b>			
4.1 Read again, review, reflect and evaluate	5.00	0.00	Strongly agree
<b>Overall</b>	4.81	0.27	Strongly agree

From Table I, the suitability of the Cp-HL process to enhance digital natives' analytical reading skills was overall the most suitable, with a mean of 4.81 and a standard deviation of 0.27.

## VII. CONCLUSIONS AND DISCUSSION

In this article, we propose the “Cloud-powered Hybrid Learning Process” (Cp-HL process) as an instructional process involving the use of cloud-powered learning tools in hybrid learning for the purpose of enhancing digital natives' analytical reading skills by providing access to a wide range of digital resources and tools in both online and on-site learning modes according to the Sloan Consortium's HL guidelines [20], which specifies the proportion of program content delivered online from 30 to 79% as the most widely accepted and popular in Thailand. The Cp-HL process to enhance digital natives' analytical reading skills. This process consisted of four major learning phases and 11 learning steps, as follows:

1) *P1: Preparation for hybrid learning:* There were two learning steps under the online learning mode: *a)* Prepare digital devices; and *b)* Connect devices to the Internet. Cloud-powered learning tools used in this process include social communication tools (e.g., Facebook®, LINE®, Telegram®, and Discord®).

2) *P2: Presentation for interactive learning:* There were five learning steps in the online learning mode: *a)* Join the online class; *b)* Explore the story; *c)* Check the story; *d)* Read together; and *e)* Personal reading log. Cloud-powered learning

tools used in this process include: i) browsing and seeking tools (e.g., Google Chrome®, Firefox®, Microsoft Edge®, and Safari®); and ii) interactive presentation tools (e.g., Zoom®, Google Meet®, Microsoft Teams®, and ClassPoint®).

3) P3: *Practice with the analytical reading*: Under the onsite learning mode, there were three learning steps: a) Join the on-site class; b) play a quiz game; and c) Share a comment involving the story using the Think-Pair-Share (TPS) technique. Cloud-powered learning tools used in this process include analytics and evaluative tools (e.g., ClassDojo®, Kahoot!®, Google Forms®, and Google Analytics®).

4) P4: *Progress reports in analytical reading skills*: There was a critical learning step under the on-site learning mode: read again, review, reflect, and evaluate the digital media. Progress reports represent the expected learning outcomes, which are digital natives' analytical reading skills, which have three components: a) identifying the intrinsic elements of the story and media symbol systems; b) analyzing the message of the story and media symbol systems; and c) responding to the story and media symbol systems. Cloud-powered learning tools used in this process include analytics and evaluative tools.

According to nine experts' evaluations of the suitability of the developed Cp-HL process, all experts agreed that the Cp-HL process was the most suitable overall for improving digital natives' analytical reading skills, with a mean of 4.81 and a standard deviation of 0.27. Based on these findings, we anticipate that our CP-HL process will be implemented in both online and onsite learning modes that integrate digital reading techniques, hybrid pedagogies, and cloud-powered learning tools to promote digital natives' analytical reading skills, which align with the characteristics of digital learners.

In the future directions of hybrid learning research [17, 32], it will still be challenging to integrate pedagogical strategies with new learning technologies to help learners of all ages gain access to learning, help people learn, and achieve the learning outcomes that society needs in lifelong learning, including those with certain barriers to learning, whether they are in formal education, non-formal education, or alternative education.

#### ACKNOWLEDGMENT

The Division of Learning Innovation and Technology, Faculty of Industrial Education and Technology, King Mongkut's University of Technology Thonburi (KMUTT) supported the research, and the research team would like to thank Dr Kridsanapong Lertbumroongchai for advice on the preparation of illustrations for the research.

#### REFERENCES

- [1] National Strategy Secretariat Office. (2018, October 8). Thailand National Strategy 2018-2037 (Summary). Office of International Affairs. Retrieved March 1, 2021, from <https://oia.coj.go.th/content/category/detail/id/8/cid/5885/iid/93993>
- [2] López-Escribano, C., Valverde-Montesino, S., & García-Ortega, V. (2021). The impact of e-book reading on young children's emergent literacy skills: An analytical review. *International Journal of Environmental Research and Public Health*, 18(12), 6510.
- [3] Jamshidifarsani, H., Garbaya, S., Lim, T., Blazevic, P., & Ritchie, J. M. (2019). Technology-based reading intervention programs for elementary grades: An analytical review. *Computers & Education*, 128, 427-451.
- [4] Prawita, W., & Prayitno, B. A. (2019). Effectiveness of a Generative Learning-Based Biology Module to Improve the Analytical Thinking Skills of the Students with High and Low Reading Motivation. *International Journal of Instruction*, 12(1), 1459-1476.
- [5] Miarsyah, M., Ristanto, R. H., Lestari, P., & Rahayu, S. (2021). Metacognitive on Pteridophyte: A Unification of Cooperative Integrated Reading and Composition and Guided Inquiry (CirGI). *International Journal of Instruction*, 14(3), 481-500.
- [6] RISTANTO, R., Rahayu, S., & Mutmainah, S. (2021). Conceptual understanding of excretory system: Implementing cooperative integrated reading and composition based on scientific approach. *Participatory Educational Research*, 8(1), 28-47.
- [7] Erhan, D. (2011). Effects of cooperative integrated reading and composition (CIRC) technique on reading-writing skills. *Educational Research and Reviews*, 6(1), 102-109.
- [8] Wong, L. W., Tan, G. W. H., Hew, J. J., Ooi, K. B., & Leong, L. Y. (2022). Mobile social media marketing: a new marketing channel among digital natives in higher education?. *Journal of Marketing for Higher Education*, 32(1), 113-137.
- [9] Kahraman, E., Gokasan, T. A., & Ozad, B. E. (2020). Usage of social networks by digital natives as a new communication platform for interpersonal communication: A study on university students in Cyprus. *Interaction Studies*, 21(3), 440-460.
- [10] Broughton, A., Daly, M., Marx, N. J., Nieuwoudt, M., le Roux, D. B., & Parry, D. A. (2019). An exploratory investigation of online and offline social behaviour among digital natives. In *Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019* (pp. 1-10).
- [11] Aristika, A., & Juandi, D. (2021). The effectiveness of hybrid learning in improving of teacher-student relationship in terms of learning motivation. *Emerging Science Journal*, 5(4), 443-456.
- [12] Daher, W., Sabbah, K., & Abuzant, M. (2021). Affective engagement of higher education students in an online course. *Emerging Science Journal*, 5(4), 545-558.
- [13] Huang, R. H., et al. (2020). Guidance on flexible learning during campus closures: Ensuring course quality of higher education in COVID-19 outbreak. *Beijing: Smart Learning Institute of Beijing Normal University*.
- [14] Youhanita, E., Eryadini, N., & Nurdiana, R. (2022). Blended Learning in Revolution Industrial for Generation Z. *INTERNATIONAL JOURNAL OF ECONOMICS, MANAGEMENT, BUSINESS, AND SOCIAL SCIENCE (IJEMBS)*, 2(2), 322-331.
- [15] Li, S., & Wang, W. (2022). Effect of blended learning on student performance in K-12 settings: A meta-analysis. *Journal of Computer Assisted Learning*, 38(5), 1254-1272.
- [16] Lambrecht, K., Sweeney, M. A., & Detweiler, J. (2019). The everywhere and nowhere skill: Sustaining the assessment of Analytical Reading as critical thinking across the curriculum. *The Journal of General Education*, 68(3-4), 169-190.
- [17] Nittayathammakul, V., Chatwattana, P., & Piriyasurawong, P. (2022). Crowd context-based learning process via IoT wearable technology to promote digital health literacy. *International Education Studies*, 15(6), 27. <https://doi.org/10.5539/ies.v15n6p27>
- [18] Triyason, T., Tassanaviboon, A., & Kanthamanon, P. (2020, July). Hybrid classroom: Designing for the new normal after COVID-19 pandemic. In *Proceedings of the 11th International Conference on Advances in Information Technology* (pp. 1-8).
- [19] Staker, H., & Horn, M. B. (2012). Classifying K-12 blended learning.
- [20] Allen, I. E., Seaman, J., & Garrett, R. (2007). *Blending in: The extent and promise of blended education in the United States*. Sloan Consortium. PO Box 1238, Newburyport, MA 01950.
- [21] Watkin, A. L., & Conway, M. (2022). Building social capital to counter polarization and extremism? A comparative analysis of tech platforms' official blog posts. *First Monday*, 27(5).

- [22] Yusoff, M. N., Dehghantanha, A., & Mahmud, R. (2017). Forensic investigation of social media and instant messaging services in Firefox OS: Facebook, Twitter, Google+, Telegram, OpenWapp, and Line as case studies. In *Contemporary digital forensic investigations of cloud and mobile applications* (pp. 41-62). Syngress.
- [23] Rathod, D. (2017). Web browser forensics: google chrome. *International Journal of Advanced Research in Computer Science*, 8(7), 896-899.
- [24] Maneewan, S., Nittayathamkul, V., & Lertyosbordin, C. (2017, March). A development of knowledge management process on cloud computing to support creative problem solving skill on studio photography for undergraduate students. In *2017 6th International Conference on Industrial Technology and Management (ICITM)* (pp. 27-31). IEEE.
- [25] Karjo, C. H., Andreani, W., Herawati, A., Ying, Y., Yasyfin, A. P., & Marie, K. (2022, April). Technological Challenges and Strategies in Implementing e-Learning in Higher Education. In *2022 10th International Conference on Information and Education Technology (ICIET)* (pp. 184-188). IEEE.
- [26] Malangen, A. (2022). Information and Communications Technology (ICT) Competency and Capability of Sauyo High School Teachers: A Basis for ICT Development Plan. Available at SSRN 4183485.
- [27] Nichols, T. P., & LeBlanc, R. J. (2020). Beyond apps: Digital literacies in a platform society. *The reading teacher*, 74(1), 103-109.
- [28] Kumar, J. A., Bervell, B., & Osman, S. (2020). Google classroom: insights from Malaysian higher education students' and instructors' experiences. *Education and information technologies*, 25(5), 4175-4195.
- [29] Li, Q., Li, Z., & Han, J. (2021). A hybrid learning pedagogy for surmounting the challenges of the COVID-19 pandemic in the performing arts education. *Education and Information Technologies*, 26(6), 7635-7655.
- [30] Zainuddin, Z. (2015). The Effect of Cooperative Integrated Reading and Composition Technique on Students Reading Descriptive Text Achievement. *English Language Teaching*, 8(5), 1-11.
- [31] Dechant, E. V. (1982). 1982: Improving the teaching of reading, Englewood Cliffs, NJ: Prentice-Hall.
- [32] Luka, I. (2023), "Implementation of a blended learning course for adult learners during the COVID-19 pandemic", *Quality Assurance in Education*, Vol. 31 No. 1, pp. 91-106. (ahead-of-print).
- [33] Padzil, M. R., Abd Karim, A., & Husnin, H. (2021). Employing DDR to Design and Develop a Flipped Classroom and Project based Learning Module to Applying Design Thinking in Design and Technology. *International Journal of Advanced Computer Science and Applications*, 12(9), 791-798.



# A Model for Detecting Fungal Diseases in Cotton Cultivation using Segmentation and Machine Learning Approaches

Odukoya O.H.<sup>1</sup>, Aina S.<sup>2</sup>, Dégbéssé F.W.<sup>3</sup>

Computer Science and Engineering Department, Obafemi Awolowo University, Ile-Ife, Nigeria<sup>1,2,3</sup>

**Abstract**— This research detailed a model for detecting fungal diseases via techniques for processing images of cotton leaves. The work allowed to develop a model based on the set of preprocessed data, to formulate the developed model, to simulate and evaluate the model. It is about detecting fungal diseases in cotton cultivation. The image data records were collected in an online data repository consisting of images of cotton leaves infected with fungal diseases and normal leaf images. In addition, other images of infected and uninfected cotton leaves were collected in cotton production fields in the Ségbana region in Benin Republic. The model was formulated based on watershed segmentation technique by applying Edge Detection algorithm and K-Means Clustering; and Support Vector Machine (SVM) for classification. The simulation was done using MATLAB with Image Processing Toolbox 9.4. The results gave an accuracy of 99.05%, specificity 90%, misclassification rate 0.95%, recall rate 99.5% and precision 99.5%. In addition, with less computational effort and in less than a minute, the best results were obtained, showing the efficiency of the image processing technique for the detection and classification of infected and uninfected leaves. It was concluded that this approach was applied to detect fungal diseases on cotton leaves to promote the production and harvest of good quality cotton and valuable cotton products.

**Keywords**—Fungal diseases; watershed segmentation; SVM; K-means; Edge Detection algorithm

## I. INTRODUCTION

Cotton is a soft fibrous substance, usually white, composed of the hairs encompassing the seeds of various tropical plants with free branches (genus *Gossypium*) of the mauve family [1][2]. The ordinary cotton name comes from the Arabic "quon" and generally refers to plants that produce spinnable (plush) fibers on their integument [3].

Cotton cultivation improved the living conditions of rural and urban populations. It contributed to the economic development of the CFA franc zone countries in Africa. Production of cotton has quadrupled in the last two decades, making the region the world's second largest cotton exporter, accounting for 15% of global exports. As a result, generating the most extensive receipts in various regional countries being the main cash crop for government revenues. Cotton cultivation has impacted the poverty rate positively by employing above two million rural households [4].

Due to fungal disease problems in cotton growing, agricultural industries face economic losses and a qualitative and quantitative decline in annual production. Nowadays,

modern cotton industries are trying to ensure the quality and safety of products to provide healthy cotton to consumers.

Africa is an agricultural zone where most of the population is based on cotton farming. Farmers have a wide selection of diversity to pick appropriate cotton crops. However, the growth of these plants for greatest yield and quality products is very technical. It can be enhanced with the support of technology. The management of cotton crops needs careful observation, particularly to treat diseases that will affect production and, consequently, the life after harvest.

A disease is an alteration of one or more physiological processes caused by irritation from some factors or agents resulting (pests) in loss of coordination in plants. Plant disease(s) is a hurtful modification from the traditional functioning of the process related to physiology in a particular plant. Plant diseases generally can be classified into three major parts: Classification based on the localization area, classification based on the occurrence, classification based on the causal agent [5].

Plant disease problems have arisen as a result of a considerable drop in the quality and quantity of agricultural output. Plant disease losses in Georgia (United States) in 2007 were estimated to be over \$539.74 million. Approximately \$185 million of this total was spent on disease control, with the remainder being the result of disease damage [6]. Several pests, including bacteria, aphids, insects, fungi, and others, cause serious plant diseases, but fungi are responsible for the majority of global losses [7].

FUNGI are microorganisms that exist in the planet and are responsible for nearly 75% of all diseases. They are threadlike organisms known as HYPHAE. Anthracnose, Ascochyta blight, Black root rot, Boll rot, Charcoal rot, Leaf spot, Escobilla, Fusarium wilt, Lint contamination Rust, Phymatotrichum root rot, Powdery mildew, Sclerotium stem and root rot or southern blight, Stem canker, Seedling disease complex, and Verticillium wilt are some of the diseases that can affect your plants.

Observations by experts with the naked eye are the most available method for detecting and identifying diseases in practice. However, this requires the constant supervision of professionals, which may be unaffordable on large farms. Furthermore, in certain poor nations, farmers must travel considerable distances to reach specialists, making consultants prohibitively expensive and time-consuming.

The automatic diagnosis of plant diseases is a crucial area of research. It could be useful for large-scale surveillance cultures, allowing for the automatic detection of symptoms of the disease as soon as they emerge on the plant's leaves. As a result, seeking a rapid, automatic, less expensive, and more accurate method will be a distinct advantage for the quality and development of agricultural products and derivatives, particularly when concentrating on a certain plant and disease category. This paper aims to develop a model that can detect Fungal Diseases for cotton crops only based on quality in image (of cotton leaf) processing.

In summary, the paper's contributions can be summarized as follows:

- 1) Fungal diseases Image Acquisition was done using digital camera or any other device and it was loaded and saved using Matrox Imaging Library (MIL) software and semistructured interviews and questionnaires were also administered to gather the requirements of the model.
- 2) A classification model of fungal diseases detection was formulated based on the watershed segmentation technique using K-Means Clustering and Edge Detection algorithm and Support Vector Machine (SVM) was used for classification.
- 3) Simulation of the classification model of fungal diseases detection using Matlab's Image Processing Toolbox 9.4.
- 4) Assessing the performance of the model using accuracy, efficiency, and specificity as parameters.

This paper is organized as follows. Section II illustrates relevant literature on image processing based on plant diseases and specifically fungal diseases on cotton cultivation. Section III discusses the model design approach and methodologies adopted in the course of the research. Section IV discusses an overview of the results of the simulation of the model designed and analysis of the results emanating from the simulation. Section V recaps the research work and conclusion and future work is suggested.

## II. RELATED WORKS

There are numerous literature studies on plant disease detection and image processing technologies. According to Anju [8], the watershed transformation was a good tool for image segmentation in mathematical morphology. Marker-controlled segmentation was used for watershed transformation-based segmentation. A powerful tool for image segmentation, the Prewitt Edge Detection Operator was used to demonstrate another image segmentation technique that included image enhancing and noise removal strategies. The new method was evaluated and compared to the current one. However, this work has to be improved, particularly in terms of evaluating the method's performance.

Rani and Mahip [9] published an article on the use of machine learning to detect various plant diseases using various image processing techniques. Today, the technical processing of images becomes an essential technique for diagnosing different plant diseases during cultivation. Any part of the gold harvesting area can be affected by the disease. This paper focused on detecting various diseases of cotton

cultivation and classifying them. There is as much technical classification as k-Nearest Neighbor Classifier, genetic algorithm, k-means classifier, probabilistic Neural Network, , vector support artificial analysis of the machine and main components neural network, fuzzy logic. Choosing a classification was a tedious task because the quality of the result may be different for various input data. This document provided an overview of various technical classifications used to classify plant leaf diseases.

Thikjarathi and Abirami [10] proposed an Application of Image Processing in Diagnosing Guava Leaf Diseases. In this study, scaled leaf images with improved contrasts are subjected to region expanding segmentation, color transformation (YCbCr, CIELAB), and Scale Invariant Feature Transform (SIFT). Support Vector Machines (SVM) and K-nearest Neighbor (k-NN) classifiers were investigated for disease-wise classification accuracy. While both SVM and k-NN perform well, the former has a little accuracy advantage. However, in the future, this methodology for evaluating classifiers for large datasets can be improved.

Zhang et al. [11] used a natural situation to test an artificial image segmentation algorithm for damaged cotton leaves. The authors employed a segmented monotone decreasing edge composite function, energy function guidance information, the Heaviside function, and the t penalty function to create their model  $\phi(x)$ . The results indicated that a model of a cotton leaf edge profile curve could be constructed from a model of a cotton leaf coated in bare soil, straw mulching, and plastic film mulching, and that even with non-uniformity of light, the optimal edge of the ROI could be achieved. In a complex background, the model can classify cotton leaves with uneven illumination, shadow, and weed backdrop, and it is preferable to realize the perfect extraction of the blade's edge. However, this task is restricted by the length of time it takes to complete the Segmentation (problem of efficiency). Pre-treatment, feature extraction, and picture categorization are among the other image processing processes that are bypassed.

"Detection and Identification of Rice Leaf Diseases Using Multiclass SVM and Particle Swarm Optimization Technique," by [12], for example, "By combining K-means, SVM, and multiple classifications, the author presented a new approach for detecting and identifying rice leaf illnesses." The gray level co-occurrence matrix (GLCM) was utilized to extract features. An SVM classifier was used to categorize the diseases, and Particle Swarm Optimization was used to improve detection accuracy (PSO). According to testing results, the proposed methodology was 97.91 percent accurate in terms of disease identification. Furthermore, the Feed Forward (FFNN) and SVM neuron networks accounted for 77.96 percent, 85.64 percent, and 90.56 percent, respectively, of the closest neighbour (KNN) networks.

Kumari et al. [13] suggested an automatic disease detection system for three fungal diseases in cotton crops: Alternaria leaf spot fungal disease (ALSFD), Rust Folar Fungal Disease (RFFD), and Grey Mildew Cotton Disease (GMCD). For disease segmentation on cotton leaves, the k-means clustering approach was utilized, and the results were

transferred for disease categorization to Artificial Neural Network (ANN) and Support Vector Machine (SVM) classifiers. On a Cotton plant leaf, Khalmar and Khan [14] worked on automatic early leaf spot disease segmentation. Image processing and machine learning are used in this project. The image is captured with a digital camera, and the contaminated region is segmented using a k-means clustering technique after pre-processing. The study only discovered bacterial leaf infections and did not categorize the leaves of cotton plants.

Carderia et al. [15] proposed that Deep Learning Techniques be used to identify Cotton leaf defects. This study proposes a deep learning-based method for cotton leaf screening in order to monitor the health of the leaves. In this work segmentation was not performed on the leaves, features of the leaves were identified and used for classification.

Statement of Problem Existing techniques used in detecting diseases in cotton cultivation suffer several setbacks

such as: a confusion in detection of two different diseases but having the same apparent symptoms, stages of disease detection are often incomplete and lack performance. As such, this leads to improper management of the cotton crops and causes low productivity. Therefore, there is a need for a classification model for diseases detection that will focus on specific class of the diseases (fungal diseases) to improve cotton cultivation; To fill this gap, the proposed study will provide a model that can accurately and automatically provide the right detection for fungal diseases in cotton cultivation. The outcome of this research will ensure the quality of cotton and provide healthy cotton to the consumers.

### III. PROPOSED MODEL

This section describes the developed model. This proposed model architecture depicted in Fig. 1 is an improvement of Zhang et al. [11] automation of Segmentation of infected cotton leaves under a natural environment.

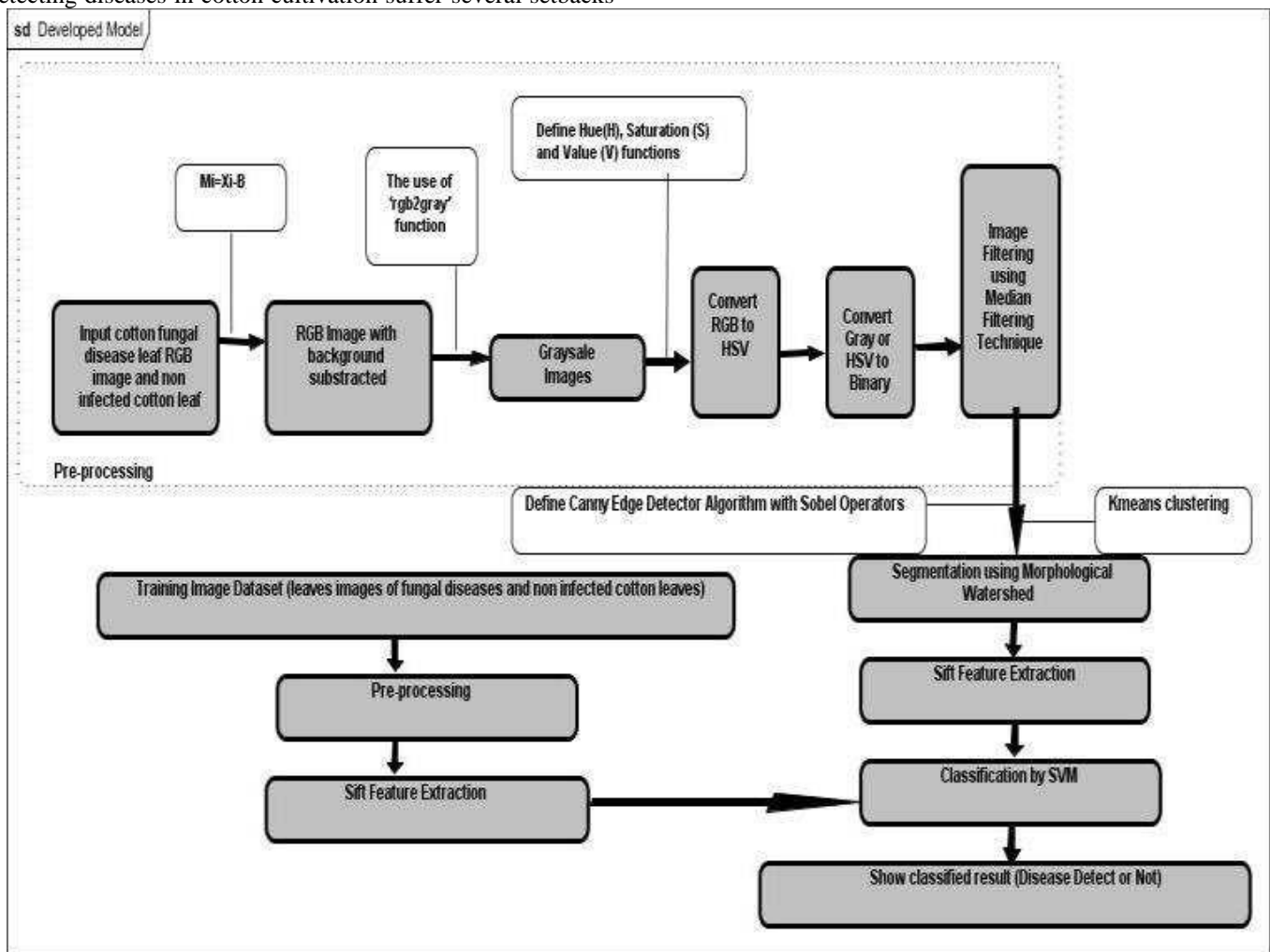


Fig. 1. Developed model.

#### A. Steps for Pre-processing

The preprocessing steps are Input picture; Subtraction of background; Converting RGB image to grayscale and HSV

image; Converting grayscale image to binary image; Filtering. Fig. 2 shows the different stages of image preprocessing.

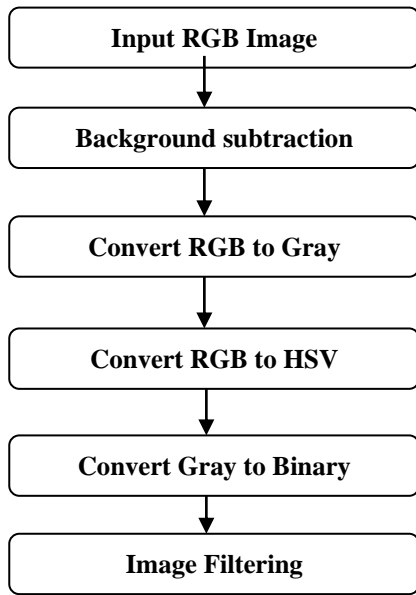


Fig. 2. Architecture flow.

### B. Development of a Formula for RGB to Grayscale

The formula of RGB to grayscale is generally expressed with a small variation of the coefficients of R, G, and B according to many papers including:

- i.  $I=0.2989R+0.5870G+0.1140B$  (1)  
according to Padmarathi and Thangadurai [16]
- ii.  $I=0.299R+0.5870G+0.1140B$  (2)  
according to Saravanan [17]
- iii.  $I= 0.3R+0.59G+0.11B$  (3)  
according to Kanan and Cottrell [18]
- iv.  $I=0.21R+0.71G+0.07B$  (4)  
according to Samuel et al. [19]

From i, ii, iii and iv luminance,

$$I=(r\%R) +(g\%G) +(b\%B) \quad (5)$$

Where

- (a) Red has contributed r%, Green has contributed g% and Blue has contributed b%;
- (b)  $r\%+g\%+b\%=100\%$  (6)

### C. Binary Image

A Binary Picture is a numeric picture. It is also considered as bi-level or two level because each pixel of this image can require a single bit 0 or 1. Generally, 0 shows the black color, and 1 shows the white color, which has two assigned pixel values. In the numeric treatment of a picture, a binary picture is used like masks or as the result of some frequent operations such as Segmentation, thresholding, and dithering, etc. Fig. 3 illustrates an example of transforming an RGB image to a grayscale image and a binary image. Fig. 4 also shows another example of transforming an RGB image to grayscale, binary and HSV image.

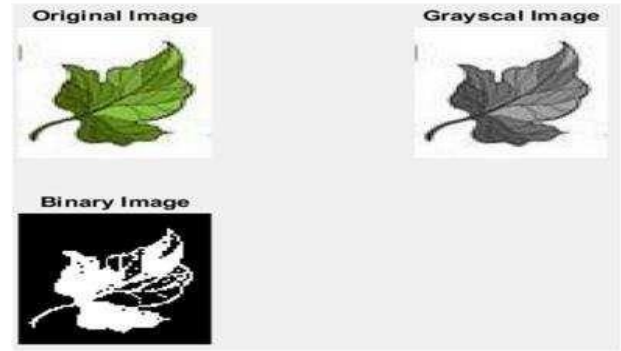


Fig. 3. Illustrates an example of converting an RGB image to a grayscale image and a binary image.

### D. Characteristic of Hue, Saturation, and Value(HSV)

$$H = \begin{cases} 0 & \text{if } \max = \min \\ \frac{(60^\circ \times g - b + 360^\circ)}{\max - \min} \text{ mod } 360^\circ & \text{if } \max = r \\ \frac{(60^\circ \times b - r + 120^\circ)}{\max - \min} & \text{if } \max = g \\ \frac{(60^\circ \times r - g + 240^\circ)}{\max - \min} & \text{if } \max = b \end{cases} \quad (7)$$

$$S = \begin{cases} 0 & \text{if } \max = 0 \\ \frac{\max - \min}{\max} = 1 - \frac{\min}{\max} & \text{otherwise} \end{cases} \quad (8)$$

$$V = \text{Max} = \frac{1(r+g+b)}{3} \quad (9)$$

Where r, g, and b stands for red, green and blue normalized in value [0, 1] [20].

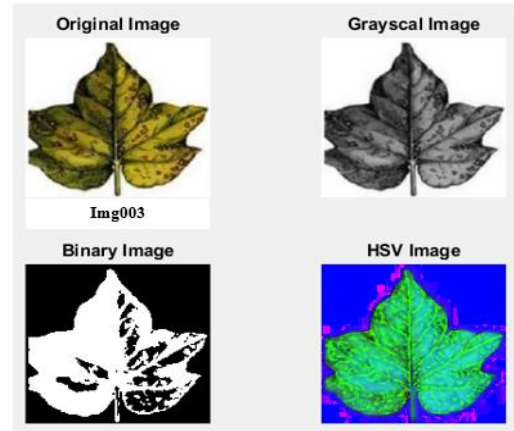


Fig. 4. illustrates an example of converting an RGB image to a grayscale image a binary image, and an HSV image.

### E. Image Filtering

In image pre-processing, filtering is an important step to use to reduce image noise and improve the visual quality of an image. Basically, image filtering is useful in various applications such as smoothing, sharpening, noise suppression, and edge detection. There are several types of image filtering techniques, such as Laplacian filtering, low-pass filtering, high-pass filtering, and so on. However, in this study, the median filtering technique was used. The image

filtering technique is used in an image pre-processing step to reduce noise to improve the result of subsequent processing such as an edge detection example of an image. The median filter is called a nonlinear digital filtering process in image processing. The median filtering technique is used in the image to suppress the noise of the image. Fig. 5 shows an example of obtaining a filtered image from an original image.

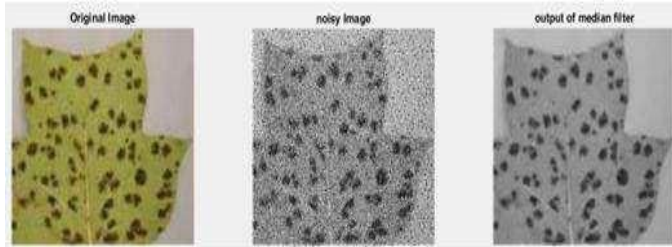


Fig. 5. Obtaining a filtered image from an original image.

**K-means Clustering:** The K-mean is frequently used to define the natural grouping of pixels in a photograph. It is a simple and quick method that attracts a significant number of consumers. For the purpose of creating vector spaces, a clustering approach is useful. The objects are grouped together at the centroids. With luminosity layer 'L\*', chromaticity-layer 'a\*' indicating where the color falls along the red-green axis, and chromaticity-layer 'b\*' representing where the color falls along the blue-yellow axis, K-Means clustering uses the L\*a\*b color format. The 'a\*' and 'b\*' layers contain all color information. The generalized pseudocode of the traditional k-means algorithm and the traditional k-means method according to Oyelade et al. [21] are shown in Algorithm 1 and Algorithm 2, respectively.

---

**Algorithm 1:** Generalized pseudocode of the traditional k-means

1. Accept the number of clusters to group data into and the clusters as input values.
  2. Initialize the first k clusters
    - 2.1 Take the first k instances or
    - 2.2 Take Random Sampling of k elements
  3. Calculate the arithmetic mean of each cluster
  4. K-means assigns each record in the dataset to only one of the initial clusters.
    - 4.1 Each record is assigned to the nearest cluster using a measure of Distance (e.g Euclidean distance).
  5. K-means re-assigns each record in the dataset to the most similar cluster and re-calculates the arithmetic mean of all the clusters in the dataset.
- 

---

**Algorithm 2:** Traditional k-means algorithm

**INPUT:** Dataset of n data points  $x_1, x_2, \dots, x_n$

**OUTPUT:**  $m_j (j = 1, 2, \dots, k)$  = Cluster centroids

MSE (Mean Square Error) = Large Number

Select Initial Cluster centroid  $m_j$ ;

While (MSE < OldMSE) do

OldMSE = MSE;

MSE1 = 0;

For (j = 1 to k)

$m_j = 0; n_j = 0;$

endfor

For (j = 1 to k)

Compute Squared Euclidean Distance  $d^2(x_i, m_j)$ ;

endfor

For (j = 1 to k)

Find the closest centroid  $m_j$  to  $x_i$ ;

$m_j = m_j + x_i; n_j = n_j + 1;$

$MSE1 = MSE1 + d^2(x_i, m_j)$ ;

endfor

For (j = 1 to k)

$m_j = m_j + x_i;$

$n_j = n_j + 1;$

$MSE1 = MSE1 + d^2(x_i, m_j)$ ;

$n_j = \max(n_j, 1)$ ;

$m_j = m_j / n_j;$

endfor

MSE = MSE1;

---

**F. Edge Detection**

The identification of edge is employed in image analysis to determine the region's boundaries. In human vision, and probably in many other biological vision systems as well, edges and contours play a significant role. It is not just the sides that catch the eye; a few essential lines can frequently be used to describe or reconstruct an entire figure. Edges are considerable local changes of intensity in a picture. The reasons for intensity include events of geometry (orientation of surface (boundary) discontinuities, discontinuities of depth, discontinuities of color and texture) and non-geometric events (changes of illumination, specularities, inter-reflections of shadows).

There are many techniques for the recognition of edges on a picture. In this study, an appropriate estimate of the first derivative determined by two operators of Sobel will be used. As the derivatives heighten noise; the smoothing aspect is an interesting feature of operators of Sobel. Firstly, the derivatives are executed with the use of gradient amplitude. For a function  $f(x, y)$ , the gradient  $f$  at the coordinate  $(x, y)$  is determined as the two-dimensional column vector. Fig. 6 shows the edge detection with the use of Sobel operators in segmentation.

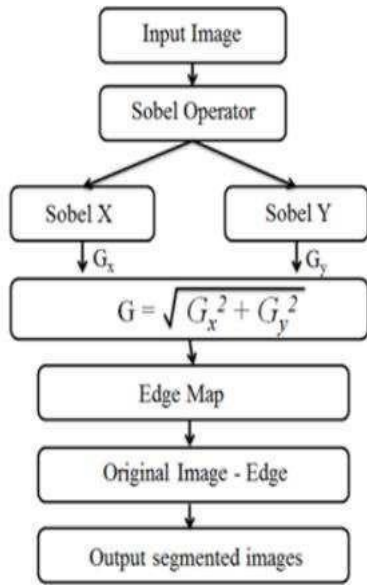


Fig. 6. The edge detection using Sobel Operators in Segmentation.

$$\Delta f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \partial f / \partial x \\ \partial f / \partial y \end{bmatrix} \quad (10)$$

$$\begin{aligned} \Delta f = \text{mag}(\Delta f) &= [G_x^2 + G_y^2]^{1/2} \\ &= \left[ \left( \partial f / \partial x \right)^2 + \left( \partial f / \partial y \right)^2 \right]^{1/2} \\ &\cong [G_x + G_y] \quad (11) \end{aligned}$$

Where

$G_x$  is the gradient along the x direction

$G_y$  is the gradient along the y direction

Sandiya and Patial [22].

### G. Image Segmentation

The process of segmenting image is breaking down a numeric image to form many segments. Segmentation is a technique for giving meaning to a picture or making an analysis more intelligible. The Segmentation of the picture is used to localize the limits and objects as curves, lines, etc. Literally, in image segmentation operation, all pixels in a picture have been assigned a label where the same label allots the same visual features. The outcomes of image segmentation are multiple segments that a set of contours extracted from the picture or collectively cover the whole image. In a region, each pixel is a computed property or similar with respect to some feature, such as color, texture, or intensity.

### H. Morphological Watershed Segmentded Images

The key to using the watershed transform for segmentation is to change your image into one with catchment basins that match to the objects you wish to recognize. Infected and non-infected leaves are shown in Fig. 7 as a Morphological Watershed Segmented Image.

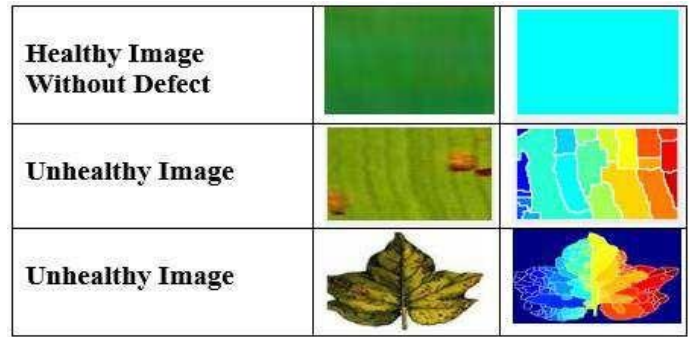


Fig. 7. Infected versus non-infected leaves in a morphological watershed image.

### I. Watershed Approach

#### 1) The Algorithm

The possibilities of the algorithm are as follows: assume that a hole is drilled at least in each regional neighborhood and that all geology is submerged from below, allowing the water to climb at a constant rate through the perforations. Flooded pixels are those that are under the water level at any given time. As the water level increases, the size of the flooded areas increases. Water will reach a level where two sites immersed in separate catchment basins will mix in the long term. When this occurs, the algorithm constructs a one-pixel thick barrier between the two sections. This overflow continues till the entire image is divided into numerous different basins separated by watershed edge lines or ridgelines.

### J. Feature Extraction: Sift-Scale Invariant Feature Transform

Any object can have various features and important details that can be accessed to produce a description. This layout is used to identify an element in an image that contains a lot of other things. The SIFT method converts a picture into a large set of local feature vectors in order to generate image characteristics. The image's size, translation, or rotation have no effect on any of these feature vectors. The SIFT algorithm uses a four-stage filtering approach to help in the extraction of these properties:

- 1) Detection of Extremes in Scale and Space
- 2) Keypoint Location
- 3) Assignment of Orientation
- 4) Descriptor of a Keypoint

### K. Development of Formula for Difference of Gaussians

Calculation  $D(x, y, \sigma)$

$$D(x, y, \sigma) \text{ if } L(x, y, \sigma) = I(x, y) * G(x, y, \sigma)$$

$$\text{Then } L(x, y, k\sigma) = I(x, y) * G(x, y, k\sigma)$$

$$\text{Therefore } D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

$$= I(x, y) * G(x, y, k\sigma) - I(x, y) * G(x, y, \sigma)$$

$$= I(x, y) * [G(x, y, k\sigma) - G(x, y, \sigma)]$$

$$\text{If } G(x, y, \sigma) = \frac{1}{2\pi e^2} e^{-\frac{(x^2+y^2)}{2e^2}} \text{ then}$$

$$D(x, y, \sigma) = I(x, y) * [G(x, y, k\sigma) - G(x, y, \sigma)]$$

$$= I(x, y) * \left[ \frac{1}{2\pi k e^2} e^{-\frac{(x^2+y^2)}{2ke^2}} - \frac{1}{2\pi e^2} e^{-\frac{(x^2+y^2)}{2e^2}} \right]$$

$$= \frac{1(xy)}{2\pi e^2} * \left[ \frac{1}{k^2} e^{-\frac{(x^2+y^2)}{2e^2}} \right]^{1/k^2} - e^{-\frac{(x^2+y^2)}{2e^2}} \right]$$

But with  $A = \frac{I(xy)}{2\pi e^2}$  and  $B = e^{-\frac{(x^2+y^2)}{2e^2}}$

$$D(x, y, \sigma) = A \times B^{\frac{(B)^{1/k-B}}{k^2}} \quad (12)$$

Where:

The difference of Gaussians calculation is  $D(x, y, \sigma)$ .

\* Denotes the convolution operator.

$I(x, y)$  is the input picture.

Then the scale space is calculated by the function:

$$L(x, y, \sigma) = I(x, y) * G(x, y, \sigma)$$

with  $G(x, y, \sigma) = \frac{1}{2\pi e^2} e^{-\frac{(x^2+y^2)}{2e^2}}$

#### L. Support Vector Machine (SVM) Classification

SVM is a technique for data analysis and recognition. It is used to do things like regression analysis and categorization. The SVM algorithm takes a set of input data and predicts classes for each individual input. SVM is a linear model-based trainer that uses extracted features to build a hyper plane that translates a piece of data into the separation between the hyperplane and the nearest training points.

### IV. RESULTS

#### a) Contrast

The intensity difference between a pixel as well as its adjacent for the entire image is returned by this function. Range=[0(size(GLCM, 1) - 1)2]. There's no difference in a continuous image.

$$Contrast = \sum_{i,j=0}^{N-1} (i, j)^2 \quad (13)$$

#### b) Energy

The GLCM Range = [0 1] returns the total of squared elements. A continual image necessitates the use of energy.

$$Energy = \sum_{i,j=0}^{N-1} C(i, j)^2 \quad (14)$$

#### c) Homogeneity

Returns a value that indicates how near an element's GLCM distribution is to the GLCM diagonal. [0 1] is the range. For a diagonal GLCM, homogeneity is 1

$$Homogeneity = \frac{\sum_{i,j=0}^{N-1} C(i, j)}{(1+(i-j))^2} \quad (15)$$

According to the results of Table I, the normal leaves of cotton have the following features: the contrast of different leaves is between 0 to 0.09; homogeneity is between 0.98 to 1.0; energy is between 0.98 to 1.0. According to the results of Table II, the disease leaves of cotton have the following features: the contrast of different leaves is between 0.1400 to 0.3; homogeneity is between 0.2000 to 0.97; energy is

between 0.3 to 0.97. The above result analysis concludes that diseased leaf image contrast is more than the normal leaf. On the other hand, diseased leaf image homogeneity and energy are less than the normal leaf. Hence, we conclude that images are defective.

TABLE I. NORMAL LEAF IMAGE ANALYSIS

Normal	Leaf1	Leaf2	Leaf3
CONTRAST	0.0898	0.0501	0.0
ENERGY	0.9879	0.9897	1.0
HOMOGENITY	0.9807	0.9759	1.0

TABLE II. DISEASED LEAF IMAGE ANALYSIS

Diseased	Leaf1	Leaf2	Leaf3
CONTRAST	0.2644	0.1628	0.1591
ENERGY	0.4889	0.5737	0.6834
HOMOGENITY	0.9725	0.9617	0.9343

#### A. Classification Results and Somecomparisons with the Existing Model

As the first manifestations and symptoms of the fungal disease appear on the leaves of the plant, detection techniques and classification are developed. The characteristics of color, texture, and form are widely used in detecting and classifying plant infections. Following segmentation, the texture, pigment, and form properties of the diseased areas are removed or separated and used as input to the SVM classifier.

According to Table III relating to the achievement of the preprocessing and image processing steps, only the image segmentation step is performed with the existing model while the preprocessing, segmentation, extraction, and classification of images are done with the developed model.

Table IV provides information on the number of infected and uninfected leaves, size, and types of images in the SVM Learning database.

TABLE III. ACHIEVEMENT OF IMAGE PREPROCESSING AND PROCESSING STEPS

Leaf Type	Number of Images	Size(pixels)	Image Type
Healthy	10	193x228	JPG
Disease	500	193x228	JPG

TABLE IV. SVM LEARNING DATABASE

S/N	Steps	Existing Model [22]	Developed Model
1	Stages in image preparation	-	+
2	Categorization of image	+	+
3	Feature Extraction	-	+
4	Classification	-	+

Achieved (+).

Not achieved (-).

B. Distribution of Images

A set of 500 images of infected cotton leaves was used to validate the detection approach of the disease. In this sample of 500 images, 300 (60%) were used for the training of the system and 200(40%) for the tests. Table V illustrates the distribution of images.

TABLE V. IMAGE DISTRIBUTION

Disease	Images	Training Sample	Testing Sample
			Diseases Detection
Fungal Disease	500	300	200
Percentage	100%	60%	40%

C. Evaluation of the Model for Fungal Diseases Detection in Cotton Cultivation

The evaluation of the proposed approach was achieved based on precision and recall, and the images are grouped in accordance with the generated results. Based on the system’s results, these images were categorized as TP (rule matched and disease present), FP (rule matched and no disease present), TN(no rule matched and no disease present), and FN (no rule matched and disease present) based on results generated by the system.

- 1) *TP (true positive)*: The infected images are predicted to have a fungal disease.
  - 2) *TN (true negative)*: Images not infected by fungal disease and are predicted to be negative of infection.
  - 3) *FP (false positives)*: No fungal diseases present in the image but predicted to be present by the system.
  - 4) *FN (false negatives)*: Presence of fungal disease in the image but the system predicted its absence.
  - 5) *GT (ground truth)*: It displays the number of comparisons made during the testing process.
  - 6) *TC (total cases)*: The total number of comparisons made during the testing process.
  - 7) *RM (result of method)*: This shows the overall number of false-positive and true positive system predictions.
  - 8) *Accuracy*: The performance of the system as a percentage of correct prediction.
  - 9) *Misclassification*: The percentage of times the system has predicted incorrectly.
  - 10) *Recall*: Represents the percentage of affirmative cases discovered by the system represented.
  - 11) *Precision*: It shows the percentage of the positive predictions made by the system.
- These results are presented in Tables VI and VII.

TABLE VI. THE CONFUSION MATRIX BASED ON FUNGAL DISEASES

Fungal Diseases	TP	TN	FP	FN
200	199	9	1	1

TABLE VII. SUMMARY OF THE RESULTS

Cases	Predicted: Negative	Predicted: Positive
Negative Cases	TN:9	FP:1
Positive Cases	FN:1	TP:199

TABLE VIII. EVALUATION OF THE PROPOSED METHODOLOGY BASED ON CLASSIFICATION RESULTS

Feature	Formula	Resulted Value	Accuracy (%)
GT	Actual Positive Images	200	-
TC	TP+TN+FP+FN	210	-
RM	FP+TP	200	-
Accuracy	(TN+TP)/TC	0.9905	99.05
Misclassification	(FP+FN)/TC	0.0095	0.95
Recall	TP/GT	0.995	99.5
Precision	TP/RM	0.995	99.5
Specificity	TN/(TN+FP)	0.90	90%

The evaluation results in Table VIII showed that the proposed technique gives 99.05% accuracy with a 0.95% misclassification rate. Furthermore, the recall rate is 99.5%, Specificity is 90% and precision is 99.5%. The system's efficiency is justified by reducing processing time and expertise costs. For example, under a 3.0 GHz Pentium IV PC with 1 GB of integrated RAM, the average processing time between preprocessing the image and the classification of the improved method was 49.6 seconds. The time required to obtain a result has been considerably reduced. This allows farmers to save time (a few seconds for a result) and resources (the cost of the expertise is considerably reduced, and it is not necessary to make an appointment with an expert in the hope of defining the quality of cotton).

Table IX represents the performance comparison between different cotton leaf disease identification and classification Systems.

TABLE IX. PERFORMANCE COMPARISON BETWEEN DIFFERENT COTTON LEAF IDENTIFICATION AND CLASSIFICATION MODELS

Reference	Technique	Disease affected Plant	Accuracy
Proposed System	Proposed Method with SVM Classifier	Cotton	99.5%
N. R. Bhimte and V. R. Thool [23]	Image Processing with SVM Classifier	Cotton	98.46%



## V. CONCLUSION AND FUTURE WORKS

In this research, cotton plant disease detection of infected leaves and classification of infected and uninfected leaves is done using image-based processing and machine learning approaches to assist farmers during their struggle against disease outbreaks by making the right decision to increase productivity and collect pure cotton. The results proved that the proposed model has the ability to accurately distinguish between infected and uninfected cotton leaves. However, the model is limited to only detecting fungal diseases in cotton leaves. For future research, this work could be extended to consider other categories of cotton plant diseases such as bacterial, seedling, and boll rots and also implement the system as a mobile and web application to eradicate the manual identification of plant defects, which has become a long and costly process.

In general, the use of image processing to detect diseases and classification does not apply to a specific area. For sustainable agriculture, this work can be applied to identify diseases or the quality of vegetables with great accuracy. This work will contribute significantly to the development of agricultural research.

## ACKNOWLEDGMENT

This Research was funded by the TEDFUND Research Fund and Africa Center of Excellence OAK-Park, Obafemi Awolowo University, Ile-Ife, Nigeria.

## REFERENCES

- [1] Wendel, J. F., and Albert, V. A. (1992). Phylogenetics of the cotton genus (*Gossypium*): character-state weighed parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. *Systematic Botany*, 115- 143.
- [2] Seelanan, T., Schnabel, A., & Wendel, J. F. (1997). Congruence and consensus in the Cotton tribe (Malvaceae). *Systematic Botany*, 259-290.
- [3] Lee J (1984). Cotton as a World Crop. *Agronomy. A series of monographs - American Society of Agronomy (USA) Journal*. 24,1-25. ISSN : 0065-4663.
- [4] Badiane, O.G., Dheineswar, G.L. (2002)"Cotton Sector Strategies in West and Central Africa(English). Policy Research working paper;no: WPS 2867. Washington, D.C. ; World Bank Group <http://documents.worldbank.org/curated/en/481531468768541415/Cotton-Sector-Strategies-in-West-and-Central-Africa>.
- [5] Kiran R. and Ujwalla G. (2014). An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques. *IOSR Journal of Computer Engineering (IOSR-JCE)* e-ISSN:2278-0661, p-ISSN: 2278-8727.16(1), 10-16.
- [6] Al-Hiary H., Bani-Ahmad S., Reyalat M., Braik M. and ALRahamneh Z. (2011). Fast and Accurate Detection and Classification of Plant Diseases. *International Journal of Compute Applications (0975 – 8887)*.17(1), 31-38.
- [7] Asher M. and Hanson L. (2006). *Fungal and Bacterial Diseases*. pp. 286-315 In Sugar Beet. A.P. Draycott, ed. Blackwell Publishing Ltd.
- [8] Anju,B. (2012)" An Improved Watershed Image Segmentation Technique using Matlab". *International Journal of Scientific and Engineering Research*. Volume3, Issue 6. ISSN:2229-5518.
- [9] Rani P, and Mahip B. (2014). Review Paper on Identification of Plant Diseases Using Image Processing Technique. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*. 3(11),3667-3669. ISSN: 2278 – 1323.
- [10] Thilagavathi, M, and Abirami, S. (2017)" Application of Image Processing in Diagnosing Guava Leaf Diseases". *International Journal of Scientific Research and Management (IJSRM)*.Volume 5, Issue 07, pp 5927-5933.
- [11] Zhang Jian-hua, Kong Fan-tao, Wu Jian-zhai, Han Shu-qing, Zhai Zhi-fen (2018). Automatic image segmentation method for cotton leaves with disease under natural environment. *Journal of Integrative Agriculture*. 17(8): 1800–1814.
- [12] Prabira, K.S., Nalini, K.B. and Amuya, K.B.(2019) " Detection and Identification of Rice leaf Disease using Multiclass SVM and Particle Swarm Optimization Technique".*International Journal of Innovative Technology and Exploring Engineering(IJITEE)*. Volume 8, Issue 6S2 pg 108-120. ISSN:2278-3075.
- [13] Kumari, C.U., Vignash, N.A., Panigraphy, A.K. ,Ramya, L. and Padma, T.(2019) "Fungal Disease in Cotton Leaf Detection and Classification using Neural Networks and Support Vector Machine, Volume 8, Issue 10. ISSN:2278-3075.
- [14] Khairnar, K. and Khan, S. (2020)" Automatic Early Leaf Spot Disease Segmentation on Cotton Plant Leaf ". *International Journal of Recent Technology and Engineering*. Volume 9, Issue 2, ISSN: 2277-3878.
- [15] Carderia, R.F., Santiago, W.E., Teruel, B. (2021) Identification of Cotton Leaf Lesions Using Deep Learning Techniques. *Sensor*, 21,3169, <https://doi.org/10.3390/s21093169>.
- [16] Padmavathi, K. and Thangadurai, K. (2016)" Implementation of RGB and Grayscale Images in Plant leaves Disease Detection a Comparative Study. *Indian Journal of Science and Technology*. Volume 9, no 6.pg 1-6. DOI:10.17485/jst/2016/v9i6/77739.
- [17] Saravanan Chandran (2010) Color Image to Grayscale Image Conversion. *Second International Conference on Computer Engineering and Applications*.(pp196-199).DOI:10.1109/ICCEA.2010.192.
- [18] Kanan C., Cottrell GW (2012)" Color-to-Gray scale: Does the method matter in Image Recognition?. *PloS ONE*. Volume 7, no 1.pp1-7.e21740. DOI:10.1371/Journal.pone.0029740
- [19] Samuel M., Melo G., Judith K. (2015). A comparative study of grayscale conversion techniques applied to SIFT descriptors. *SBC Journal on Interactive Systems*, 6(2).30-36.
- [20] Jun-Dong Chang, Shyr-Shen Yu, Hong-Hao Chen and Chwei-Shyong Tsai (2010). HSV- based Color Texture Image Classification using Wavelet Transform and Motif Patterns. *Journal of Computers*, 20(4),63-69.
- [21] Oyelade O. J, Oladipupo O. O and Obagbuwa I. C (2010). Application of k-Means Clustering algorithm for prediction of Students' Academic Performance. *(IJCSIS) International Journal of Computer Science and Information Security*,7(1),292-295.ISSN1947-5500.
- [22] Sandiya, V. and Patial, R. (2012) Sobel Edge Detection using Parallel Architecture based on FPGA. *International Journal of Applied Information Systems (IAIS)*.3(4),20-24, ISSN:22490868
- [23] N. R. Bhimte and V. R. Thool (2018) "Diseases Detection of Cotton Leaf Spot Using Image Processing and SVM Classifier," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 340–344.

# Deep Learning Models for the Detection of Monkeypox Skin Lesion on Digital Skin Images

Othman A. Alrusaini

Department of Engineering and Applied Sciences  
Applied College, Umm Al-Qura University  
Makkah 24382, Saudi Arabia

**Abstract**—The study is an investigation testing the accuracy of deep learning models in the detection of Monkeypox. The disease is relatively new and difficult for physicians to detect. Data for the skins were obtained from Google via web-scraping with Python's BeautifulSoup, SERP API, and requests libraries. The images underwent scrutiny by professional physicians to determine their validity and classification. The researcher extracted the images' features using two CNN models - GoogLeNet and ResNet50. Feature selection from the images involved conducting principal component analysis. Classification employed Support Vector Machines, ResNet50, VGG-16, SqueezeNet, and InceptionV3 models. The results showed that all the models performed relatively the same. However, the most effective model was VGG-16 (accuracy = 0.96, F1-score = 0.92). It is an affirmation of the usefulness of artificial intelligence in the detection of the Monkeypox disease. Subject to the approval of national health authorities, the technology can be used to help detect the disease faster and more conveniently. If integrated into a mobile application, it can be members of the public to self-diagnose before seeking official diagnoses from approved hospitals. The researcher recommends further research into the models and building bigger image databases that will power more reliable analyses.

**Keywords**—Monkeypox; digital skin images; artificial intelligence; deep learning; convoluted neural networks; VGG-16

## I. INTRODUCTION

In the recent past, the world has experienced a pandemic and is still recovering from its adverse effects. Unfortunately, as COVID-19 diminishes in incidence and prevalence, other infectious diseases, such as Monkeypox and Ebola, have sprung up. As of July 2022, 77 countries had reported at least one case of Monkeypox disease [1]. It raises questions on whether another pandemic is in the offing. Whether Monkeypox prevalence will worsen to become a pandemic or not, the disease has made a significant mark on life and livelihoods across the globe. While the virus is endemic to Central and West Africa [2, 3], the United States is the most affected, which has reported around 27,000 cases. The Center for Disease Control and Prevention (CDC) had to raise the alert level by declaring the disease a public health emergency [4, 5]. The connectivity the US has with other parts of the world socially and economically implies that it may be a matter of time before the disease spreads even further. The most challenging factor in curbing its spread is that the virus is relatively new, and physicians are still grappling with its signs and symptoms [5].

The use of artificial intelligence in the medical field is an ongoing experiment that has been recording milestones of success. The most recent accomplishment was diagnosing COVID-19 from chest X-ray images, as studies have registered close to 100% accuracy in their predictions [6-8]. It begs the question of whether Artificial Intelligence (AI) scientists can extrapolate this methodology and apply it to the Monkeypox scourge. According to [9], the most significant signs of the disease are evident on a patient's skin. The study reports that such patients bear a rash on their skin. It is one of the many signs an individual experiences when infected with the disease [10]. Since it is a visible mark on a patient's skin, one would argue that physicians should be able to diagnose using their highly experienced eyes. However, the biggest problem with this assumption is that most physicians encounter these cases for the first time in their careers [11]. Another problem is that no scientifically proven lab tests can accurately diagnose the disease, especially in its early stages [9].

Another problem is that the rashes exhibited by Monkeypox are almost similar to those experienced by patients suffering from measles, chickenpox, smallpox, and cowpox. One would have to compare and contrast the patients' skins to tell one from the other. Such a process requires that a physician has access to patients with all other similar diseases, which is untenable [12]. Additionally, the chances of committing errors of judgment are high. Artificial intelligence can rid the diagnosis process of these bottlenecks because of its high accuracy and proven reliability in the past. Researchers are continuously creating databases of Monkeypox and other pox images to aid in the classification and isolation of the virus to curb its spread [13]. Hence, this paper tests the accuracy of machine learning models in classifying digital skin images to detect Monkeypox. A high F1-score, accuracy rate, and convincing confusion matrices should be sufficient to provide evidence that artificial intelligence is applicable in this situation. Therefore, this study investigates and testing the accuracy of deep learning models in the detection of Monkeypox on digital skin images by using different models: Support Vector Machines, ResNet50, VGG-16, SqueezeNet, and InceptionV3.

The literature review section examines existing evidence on the topic where the researcher discusses what other studies have accomplished or failed to do so. The methodology section formulates the study's data collection, feature extraction, selection, classification, and evaluation plan. In

results section, the paper compares different models in detecting Monkeypox. Afterwards, the researcher discusses these findings alongside what other studies have reported. The conclusion section explains the implications of the research and makes recommendations based on the findings.

## II. LITERATURE REVIEW

Several studies have investigated the reliability of artificial intelligence in diagnosing Monkeypox using digital skin images. The study by [14] decries the rarity of Monkeypox as the cause of the knowledge gap, which inspired the investigation. The source employed deep machine learning techniques in sourcing, preparing, and testing the image data. Findings from the research indicated AI's precision of 0.85 and a mean accuracy score of 0.83. The confusion matrices developed in the study affirm the reliability of these tests to produce accurate results. Another study on this topic is by [15], which evaluates a modified VGG-16 model. The researchers also sourced digital images from online sources and were keen only to select those with licenses. The results from the research indicate that the modified model can detect Monkeypox with an accuracy of 0.97 in the first study and 0.88 in the second study [15]. The findings present a case for using AI techniques to diagnose potential Monkeypox patients.

Some studies have utilized transfer learning techniques for feature extraction. The study by [16] employed transfer learning and GoogLeNet deep network to handle its feature extraction procedures. The paper utilized publicly available datasets to evaluate hybrid classification algorithms. Results showed that, on average, the test accuracy was 0.99. The study by [1] examined the differences between warts caused by HPV and Monkeypox. In the investigation, the researchers used DNA mapping to determine whether an individual has Monkeypox, HPV, or is healthy. Findings established that the classification algorithm managed an F1-score of 0.99 and an average accuracy score of 0.96. Similarly, the investigation by [13] used MATLAB and TensorFlow to classify skin lesion images in the detection of Monkeypox. Moreover, the mention study was particularly unique in that it created a new mobile application that would be used to scan new digital images and report the classification results. The goal was to provide a preliminary system that people with skin anomalies can use in determining whether they have a reason to worry. The results establish an accuracy score of 0.91. Even with this accuracy, the researchers still encourage people to visit hospitals for check-ups regardless of the results from the mobile application.

Some studies first built their image databases before attempting to run the analysis. An excellent example of such a study is [17]. The paper is elaborate in its approach to classifying skin lesion images to detect Monkeypox. The researchers first developed the Monkeypox Skin Lesion Dataset to include two other pox diseases, namely Measles and Chickenpox. They sourced the images from websites, case reports, and news portals. They were careful only to include publicly accessible and non-commercial images. In selecting the experimental set-up, the study adopted the 3-fold cross-validation method. Similar to the approach by [18], the

researchers then augmented this data using various techniques to create a broader database. Augmentation enhanced the dataset's size by increasing the number of images from 228 to 3,192. Classification accuracy was 0.83 for the ResNet50 model and 0.79 for the InceptionV3 model. The VGG-16 model scored 0.81 accuracy.

## III. METHODOLOGY

This section explains the methodological steps taken by the researcher in obtaining and analyzing data. The first subsection outlines the researcher's data collection plan, which is a critical part of the project. The other subsections explain the steps taken by the study in feature extraction, feature selection, image classification, and model evaluation. The flowchart for the experiment is shown by Fig. 1.



Fig. 1. Flowchart of the experiment

### A. Data Collection

The study is a research paper testing the accuracy of machine learning models in classifying digital skin images to detect Monkeypox disease. Pictorial data on the disease is still scanty. Many studies, such as [14, 15], have constructed databases of Monkeypox images that the researcher could have used. Nevertheless, the researcher was interested in conducting primary research, and contributing to the discourse by giving an independent opinion about the viability of machine learning models in detecting the disease. However, the researcher established that there are several images on the web that could become potential candidates for this analysis. Hence, the study used a web-scraping tool to search Google for Monkeypox, Measles, Smallpox, and healthy skin images. Using the requests, SERP API, and BeautifulSoup, the researcher obtained images from the search engine. The three libraries mentioned above are not the only ones employed but are the most crucial in web scraping for images on Google [19]. While there were many other images, the study confined itself to common license images to avoid unnecessary copyright infringements.

The study also hired one expert physician to screen the data to confirm its validity. The researcher targeted to have an equal number of images for each class. By the end of this screening exercise, the number of Monkeypox, Measles, Smallpox, and healthy skin images was 200 (50 for each class). While some classes had more images than others, the researcher only picked 50 so that he could maintain an equal distribution of items across the classes. Most images needed further processing, which involved cropping and removing any marks that anyone could use to identify the person in the image. Furthermore, the data was augmented by adjusting brightness, rotating, modifying sharpness, zooming, and shearing. In the end, the five augmentation techniques produced 1000 images out of the 200 original images [14]. Fig. 2 shows examples of Monkeypox, Measles, Smallpox, and healthy skin images.

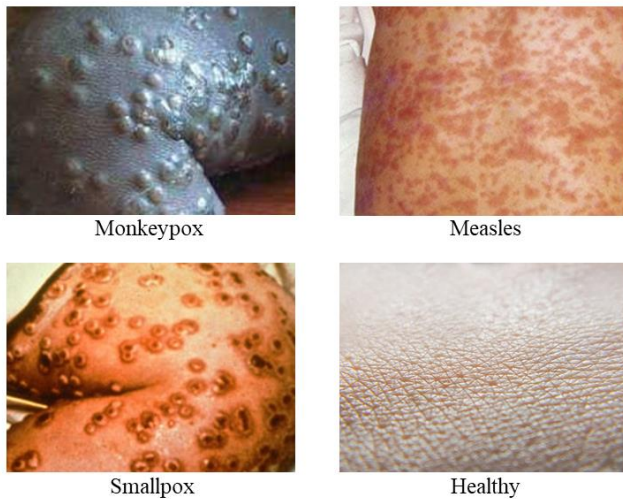


Fig. 2. Example skin images of monkeypox, measles, smallpox, and healthy cases

### B. Feature Extraction

The next step succeeding the pre-processing stage is feature extraction. It is important in machine learning because it significantly reduces the noise during analysis [20, 21]. In this case, only the most critical features will end up as potential input items. The accuracy of learned models may be improved by extracting features from input data. The process eliminates duplicate data to diminish the objects' dimensions [22]. The outcome is that the time a model takes to learn the data substantially reduces. It involves technical procedures in the background, such as combinations and transformations [23]. This study considered CNN techniques in extracting features from the images. While CNN models are primarily used in classification, they also have powerful extracting capabilities. The specific CNN models used to extract features in this study were GoogLeNet and ResNet50. GoogLeNet is a 22-layer CNN with a pretrained version that can readily images into more than 1000 object categories. ResNet50 is also a CNN but has 50 layers to it and is more computational stronger than GoogLeNet [24]. The two were used in conjunction with each other to extract features from the image dataset.

### C. Feature Selection

The researcher settled for the principal component analysis technique as the feature selection method. It involves obtaining the eigenvectors of a covariance matrix with the most significant eigenvalues and then using their respective eigenvectors [25]. These values then form the basis of feature selection. In this study, the researcher squared the standard deviations of the variables to obtain their variances. Variables with the highest variances were then retained, while those with lower variances were discarded. The PCA is used as a matrix dimension reducer because it examines components and selects only those that meet a specific criterion [25]. This procedure makes the modeling process more efficient by reducing the time needed to run the machine learning procedures due to many unimportant variables.

### D. Classification

The classification phase is the stage at which the detection actually occurs. Several models can be used to classify the image data into different groups depending on their features. The researcher settled five models, namely Support Vector Machines (SVM), ResNet50, SqueezeNet, VGG-16, and InceptionV3. SVM is a deep learning technique that adopts a supervised learning approach with associated learning algorithms to classify or regress items [26]. It creates a hyperplane, which is also the decision boundary and the basis of the classification [27]. ResNet50 is a 50-layered CNN [17]. It is a robust algorithm for image classification, as it won the ImageNet challenge in 2015. VGG-16 is also a convolutional neural network that is 16 layers deep [23]. Its strength is in its implementation in that it is simple to use. InceptionV3 is a convolutional neural network primarily used in image classification and object detection [18]. It is highly applicable and is one of the modules used in GoogLeNet. SqueezeNet is an 18-layer CNN mostly used for computer vision. The model was developed and is maintained by researcher resident at the University of California. A pretrained SqueezeNet model is capable of classifying several categories of items including most common objects and animals.

### E. Evaluation

In evaluating the models, the researcher's interest is in their accuracy, precision, recall, and F1-scores. Accuracy is the ratio of all correct predictions against all possible predictions. The precision metric measures the ratio of truly positive predictions against the number of the actual positives in a dataset. The recall metric is almost similar to precision, as it measures the proportion of cases predicted as positive that are, in fact, positives. The F1-score is the harmonic mean of the true positive rate (recall) and the precision. Hence, this study will consider the accuracy and F1-score as the most critical metric in determining the reliability of the models in detecting Monkeypox from digital skin images. According to [28], the F1-score should be at least 0.90 for a machine learning modeling process to be effective in carrying out predictive analysis. The F1-score is used as the basis of determination because it combines the usefulness of two competing metrics (recall and precision). Nevertheless, all other metrics will be reported and analyzed. The following equations were utilized to compute these metrics [15]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Where TP is the true positives, TN is the true negatives, FP is the false positives, and FN is the false negatives.

## IV. RESULTS

This section presents and analyzes the findings in training and testing the five models to detect Monkeypox disease from digital skin images dataabse explained in detail in the data

collection section. The models used in this analysis are SVM, ResNet50, VGG-16, SqueezeNet, and InceptionV3. The researcher considered a 5-fold cross-validation in enhancing the models' predictive capability since the data was not expansive enough to utilize train, validation, and splitting operations. Results in Table I show the mean metrics from the five folds. The model with the highest quality score across the five folds is VGG-16, which obtained a mean accuracy of 0.96 and a mean F1-score of 0.92. The overall performance of the five models (SVM, ResNet50, SqueezeNet, VGG-16, and InceptionV3) was reasonably close to each other. The least effective classifier model was SqueezeNet, which had an accuracy of 0.86 and an F1-score of 0.74. Other metrics

(precision, recall, and individual F1-scores) are also presented in the Table I below.

In the analysis involved attempting to make predictions using the models examined above. The outcome is presented as confusion matrices for the most effective model (VGG-16) as shown in Fig. 3 below. It is also noteworthy that there were no instances that the model reported a healthy person as being infected by Monkeypox. This fact is exemplified by healthy skin row of the confusion matrices below, where no healthy skin was detected as having Monkeypox. It adds to its reliability as a detector for the disease. Appendix A shows the matrices for the other models.

TABLE I. MODEL RESULTS ACROSS 5-FOLD CROSS-VALIDATION INSTANCES

Models	Metrics	Fold #1	Fold #2	Fold #3	Fold #4	Fold #5	Mean
SVM	Precision	0.79	0.85	0.74	0.77	0.80	0.79
	Recall	0.84	0.82	0.70	0.88	0.78	0.80
	F1-score	0.82	0.84	0.72	0.82	0.79	0.80
	Accuracy	0.91	0.92	0.87	0.91	0.90	0.90
ResNet50	Precision	0.79	0.82	0.72	0.77	0.81	0.78
	Recall	0.88	0.84	0.76	0.80	0.88	0.83
	F1-score	0.83	0.83	0.74	0.78	0.85	0.81
	Accuracy	0.91	0.92	0.87	0.89	0.92	0.90
SqueezeNet	Precision	0.56	0.83	0.75	0.75	0.68	0.71
	Recall	0.68	0.80	0.88	0.88	0.64	0.78
	F1-score	0.61	0.82	0.81	0.81	0.66	0.74
	Accuracy	0.79	0.91	0.90	0.90	0.84	0.86
VGG-16	Precision	0.92	0.92	0.92	0.94	0.90	0.92
	Recall	0.92	0.88	0.88	0.96	0.92	0.91
	F1-score	0.92	0.90	0.90	0.95	0.91	0.92
	Accuracy	0.96	0.95	0.95	0.98	0.96	0.96
InceptionV3	Precision	0.77	0.78	0.76	0.78	0.81	0.78
	Recall	0.74	0.72	0.84	0.92	0.68	0.78
	F1-score	0.76	0.75	0.80	0.84	0.74	0.78
	Accuracy	0.88	0.88	0.90	0.92	0.88	0.89

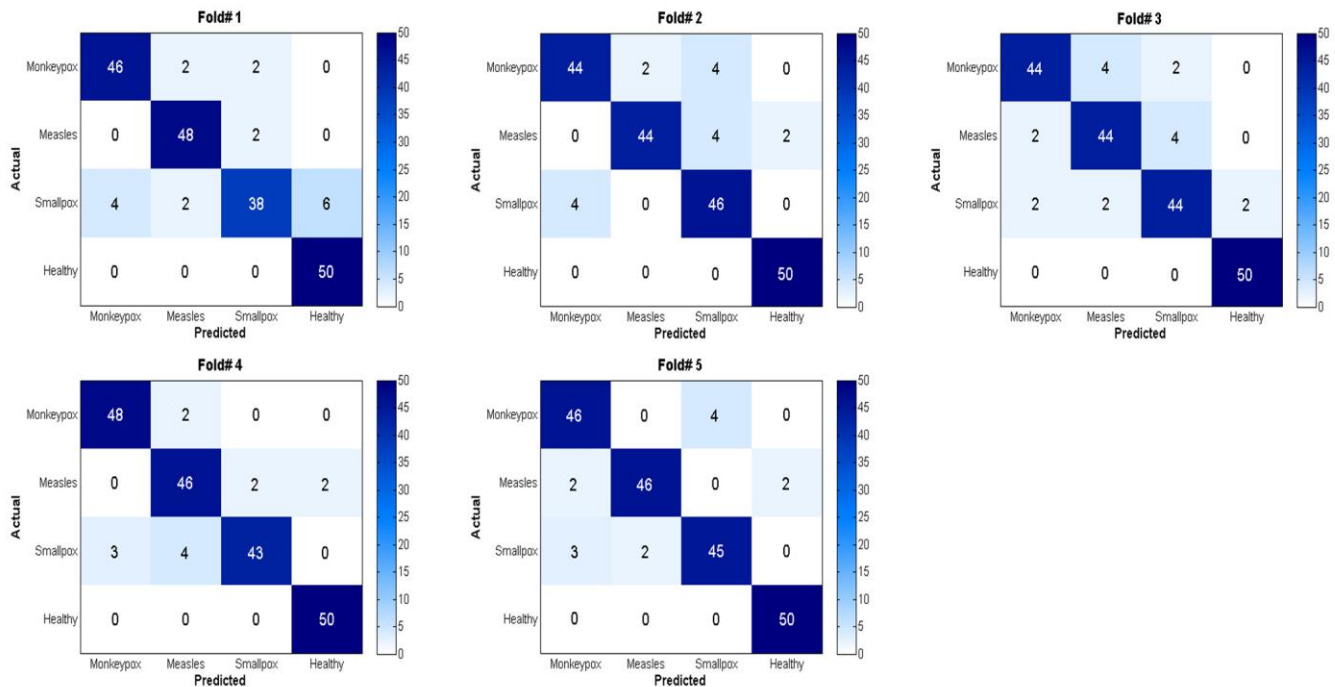


Fig. 3. Confusion matrices for VGG-16 model 5-Fold instances.

Table II below summarizes the computation of true and false positives and negatives. It shows that the model was consistent in minimizing false negatives and positives. True positives are the images correctly identified as having Monkeypox, while true negatives are the images correctly identified as having no Monkeypox. On the other hand, false positives are the images wrongly identified as having Monkeypox, while false negatives are the images wrongly identified as not infected by Monkeypox. The information is organization for each of the five cross-folds. The researcher has done the analysis using 200 images in each fold. The analysis in Table II shows that false positives and false negatives were minimal. These metrics contribute largely to the computation of recall and precision scores. For this reason, the model achieved high recall (0.91) and precision (0.92) values.

TABLE II. TRUE, FALSE POSITIVES AND NEGATIVES ANALYSIS

	FOLD #1	FOLD #2	FOLD #3	FOLD #4	FOLD #5	Sum
TP	46	44	44	48	46	228
TN	146	146	146	147	145	730
FP	4	4	4	3	5	20
FN	4	6	6	2	4	22
TOTAL	200	200	200	200	200	1000

### V. DISCUSSION

The study has established that the selected models are reasonably effective in detecting Monkeypox from digital skin images. However, the model with the highest metrics is VGG-16. It is an affirmation of the findings made in [15], where the

researchers found the VGG-16 model highly effective in detecting the disease by obtaining an F1-score of 0.97, higher than the current study’s findings. This study adds to their findings by comparing the model with similar CNN and SVM models to ascertain that the selected model gives good results. Nevertheless, there are remarkable differences between the studies. The cited source used only the VGG-16 model, while the one in focus applied five models. The approach taken by this investigation is similar to what [17] undertook. The researchers in the cited study compared VGG-16, ResNet50, InceptionV3, and Ensemble. The researchers in [17] found that ResNet50 is a better model for detecting Monkeypox because of its high F1-score (0.84) and accuracy (0.83) scores. Nevertheless, the same study found that VGG-16 is slightly lower than ResNet50 in its prediction accuracy, as it scored an F1-score of 0.83 and accuracy of 0.81. The data used by the study is dissimilar, and it may be the cause of the slight differences in the outcomes.

The researcher in this study opted for the multi-label approach because it is sometimes not enough to distinguish Monkeypox from healthy skin. Most people looking to determine their Monkeypox status usually suspect that they may have the disease because of the changes in their skin. Hence, it is important to differentiate it from other similar diseases that manifest as skin lesions. While this study dealt with four labels, other studies have dealt with even more. The investigation by [14] worked with six labels, namely chickenpox, cowpox, healthy, measles, and Monkeypox. The study also conducted 5-fold cross-validation to classify the digital skin images. It also engaged several other CNN models, which were ResNet50, InceptionV3, DenseNet121, MnasNet-A1, MobileNet-V2, ShuffleNet-V2, and SqueezeNet. The point of coincidence between [14] and the current study is that they both used ResNet50, InceptionV3,

and SqueezeNet. ShuffleNet-V2 was the most effective model, scoring the highest accuracy (0.79) and F1-score (0.67) metrics. However, the cited study did not model using SVM or VGG-16.

The ‘many models’ approach has also been evident in detecting other diseases aside from Monkeypox. The benefit of using these many models is that it allows the researcher to compare them and establish which is the most effective [6]. The paper established in the confusion matrix that the VGG-16 network did not report any false positives on persons with healthy skin, which would suggest that persons with healthy skins had Monkeypox. The author in [14] also established similar findings. In all the folds the researchers ran, there was no instance of healthy skin detected as having Monkeypox. The authors in [6] and [14] suggest that healthy skin differs significantly from that which has contracted Monkeypox. Hence, so long as one does not spot any lesions, chances are that they are safe from the disease. The lack of enough skin

images decried in [14, 15] seems to have affected accuracy scores obtained in this study. Obtaining and processing Monkeypox images may be difficult currently because of their rarity. Some of the images on Google may not be of the disease but some websites post them as Monkeypox. It is crucial that researchers hire a microbiologist to examine the skin images before using them in model training and prediction. This approach was used in [14] in creating a Monkeypox skin image database. The researcher in this investigation also shared the results with the consulted microbiologist to help in propagating the news about the technology in the profession.

Based on F1 scores, Table III below compares the performance of different deep learning models from the articles used in this paper. The table also shows the number of classifications that the model deals with, which in turn affects the accuracy of the results.

TABLE III. COMPARATIVE RESULTS WITH OTHER STUDIES

Publication	Model									Number of classifications
	VGG-16	ResNet50	InceptionV3	SVM	SqueezeNet	DenseNet121	GoogLeNet	MobileNet-V2	ShuffleNet-V2	
[13]								0.90		2
[14]		0.55	0.61			0.61		0.67	0.67	6
[15]	0.97									4
[16]		0.72					0.74			2
[17]	0.83	0.84	0.78							2
[18]	0.92	0.96					0.96	0.98	0.92	2
[23]	0.79	0.83	0.82			0.82		0.81		4
This study	0.92	0.81	0.78	0.80	0.74					4

## VI. CONCLUSION AND FUTURE RESEARCH

The current research is testing the accuracy of machine-learning models in detecting Monkeypox from digital skin images. Its findings have established that the consulted models manifested an almost similar performance. However, the outstanding model was VGG-16, whose accuracy and F1-scores were significantly higher than the rest. The detection of healthy skin was remarkably accurate because none of the healthy skin was classified as having Monkeypox. AI can be a reliable tool for physicians to differentiate between healthy skin and skin infected with the disease. Nevertheless, the accuracy in telling the difference in lesions caused by Monkeypox, measles, and smallpox still needs further analysis. The study’s F1-score of 0.92 meets the threshold of 0.90 proposed in [28], which is a vindication of the usefulness of AI in predicting the Monkeypox disease from skin images. While the achieved scores are reasonably high to suggest

proper classification, the researcher does not recommend that physicians use the technology until national health regulatory bodies further affirm the results. Other researchers should consider creating bigger databases that, when augmented, will validate the findings established in this investigation. The focus of future studies should only compare Monkeypox skin with diseases that cause lesions, such as Chickenpox, Measles, Smallpox, and Cowpox. With more research modeling the detection of Monkeypox, the researcher believes that artificial intelligence will add value to the detection process by making it quicker and more convenient. Once one has been tested, they can seek medical help and avoid contact with healthy individuals. If integrated into a mobile application, the technology can help in the detection of the disease in remote places where health facilities are fairly distant. It also provides a basis for additional research into the use of artificial intelligence in the detection of Monkeypox.

## REFERENCES

- [1] T. B. Alakus and M. Baykara, "Comparison of Monkeypox and Wart DNA Sequences with Deep Learning Model," *Applied Sciences*, vol. 12, no. 10, p. 10216, 2022.
- [2] G. H. Hans, D. Wildemeersch, and I. Meeus, "Integrated Analgesic Care in the Current Human Monkeypox Outbreak: Perspectives on an Integrated and Holistic Approach Combining Old Allies with Innovative Technologies," *Medicina*, vol. 58, no. 10, p. 1454, 2022.
- [3] E. Gomez-Lucia, "Monkeypox: Some Keys to Understand This Emerging Disease," *Animals*, vol. 12, no. 17, p. 2190, 2022.
- [4] R. J. Fischer, S. Gallogly, J. E. Schulz, N. van Doremalen, V. Munster, and S. Das, "Evaluation of Five Buffers for Inactivation of Monkeypox Virus and Feasibility of Virus Detection Using the Panther Fusion® Open Access System," *Viruses*, vol. 14, no. 10, p. 2227, 2022.
- [5] H. D. Larkin, "FDA authorizes intradermal vaccine, streamlines rules to increase monkeypox treatment access," *Jama*, vol. 328, no. 9, pp. 819-819, 2022.
- [6] F. Demir, "DeepCoroNet: A deep LSTM approach for automated detection of COVID-19 cases from chest X-ray images," *Applied Soft Computing*, vol. 103, no. 1, p. 107160, 2021.
- [7] N. W. S. Saraswati, N. W. Wardani, and I. G. A. A. D. Indradewi, "Detection of Covid Chest X-Ray using Wavelet and Support Vector Machines," *Int. J. Eng. Emerg. Technol*, vol. 5, no. 2, pp. 116-121, 2020.
- [8] O. A. Alrusaini, "COVID-19 Detection from X-Ray Images using Convolved Neural Networks: A Literature Review," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 3, pp. 78-88, 2022.
- [9] M. Jeyaraman et al., "Monkeypox: An Emerging Global Public Health Emergency," *Life*, vol. 12, no. 10, p. 1590, 2022.
- [10] E. J. Tarín-Vicente et al., "Clinical presentation and virological assessment of confirmed human monkeypox virus cases in Spain: a prospective observational cohort study," *The Lancet*, vol. 400, no. 10353, pp. 661-669, 2022.
- [11] I. Ilic, I. Zivanovic Macuzic, and M. Ilic, "Global Outbreak of Human Monkeypox in 2022: Update of Epidemiology," *Tropical Medicine and Infectious Disease*, vol. 7, no. 10, p. 264, 2022.
- [12] M. Patel, M. Surti, and M. Adnan, "Artificial intelligence (AI) in Monkeypox infection prevention," *Journal of Biomolecular Structure and Dynamics*, vol. 1, no. 1, pp. 1-5, 2022.
- [13] V. H. Sahin, I. Oztel, and G. Yolcu Oztel, "Human Monkeypox Classification from Skin Lesion Images with Deep Pre-trained Network using Mobile Application," *Journal of Medical Systems*, vol. 46, no. 11, pp. 1-10, 2022.
- [14] T. Islam, M. A. Hussain, F. U. H. Chowdhury, and B. R. Islam, "Can artificial intelligence detect Monkeypox from digital skin images?," *bioRxiv*, vol. 1, no. 1, pp. 1-7, 2022.
- [15] M. M. Ahsan, M. R. Uddin, M. Farjana, A. N. Sakib, K. A. Momin, and S. A. Luna, "Image Data collection and implementation of deep learning-based model in detecting Monkeypox disease using modified VGG16," *arXiv preprint arXiv:2206.01862*, vol. 1, no. 1, p. 1, 2022.
- [16] A. A. Abdelhamid et al., "Classification of Monkeypox Images Based on Transfer Learning and the Al-Biruni Earth Radius Optimization Algorithm," *Mathematics*, vol. 10, no. 19, p. 127, 2022.
- [17] S. N. Ali et al., "Monkeypox skin lesion detection using deep learning models: A feasibility study," *arXiv preprint arXiv:2207.03342*, vol. 1, no. 1, p. 1, 2022.
- [18] K. D. Akin, C. Gurkan, A. Budak, and H. Karatas, "Classification of Monkeypox Skin Lesion using the Explainable Artificial Intelligence Assisted Convolutional Neural Networks," *Avrupa Bilim ve Teknoloji Dergisi*, vol. 40, no. 1, pp. 106-110, 2022.
- [19] A. Rahmatulloh and R. Gunawan, "Web Scraping with HTML DOM Method for Data Collection of Scientific Articles from Google Scholar," *Indonesian Journal of Information Systems*, vol. 2, no. 2, pp. 95-104, 2020.
- [20] X. Zhang, W. Yang, X. Tang, and J. Liu, "A fast learning method for accurate and robust lane detection using two-stage feature extraction with YOLO v3," *Sensors*, vol. 18, no. 12, p. 4308, 2018.
- [21] M. Nishio, M. Nishio, N. Jimbo, and K. Nakane, "Homology-based image processing for automatic classification of histopathological images of lung tissue," *Cancers*, vol. 13, no. 6, p. 1192, 2021.
- [22] K. Zhou, M. Zhang, H. Wang, and J. Tan, "Ship Detection in SAR Images Based on Multi-Scale Feature Extraction and Adaptive Feature Fusion," *Remote Sensing*, vol. 14, no. 3, p. 755, 2022.
- [23] C. Sitaula and T. B. Shahi, "Monkeypox virus detection using pre-trained deep learning-based approaches," *Journal of Medical Systems*, vol. 4, no. 11, pp. 1-9, 2022.
- [24] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, "Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery," *Remote Sensing*, vol. 10, no. 7, p. 1119, 2018.
- [25] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: A review," *Egyptian Informatics Journal*, vol. 19, no. 3, pp. 179-189, 2018.
- [26] Y. Chen, W. Huang, L. Nguyen, and T. W. Weng, "On the Equivalence between Neural Network and Support Vector Machine," *Advances in Neural Information Processing Systems*, vol. 34, no. 1, pp. 23478-23490, 2021.
- [27] Y. Guo, Z. Zhang, and F. Tang, "Feature selection with kernelized multi-class support vector machine," *Pattern Recognition*, vol. 117, no. 1, p. 107988, 2021.
- [28] Z. Voulgaris, *Julia for Machine Learning*. Technics Publications, 2020.

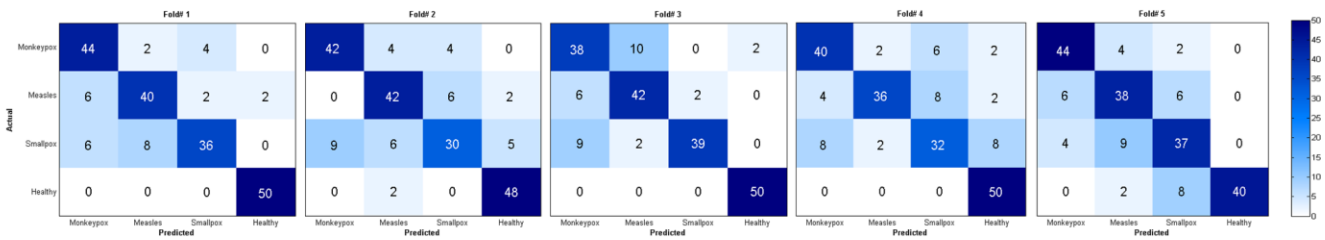


Appendix A. Confusion Matrices for Other Models

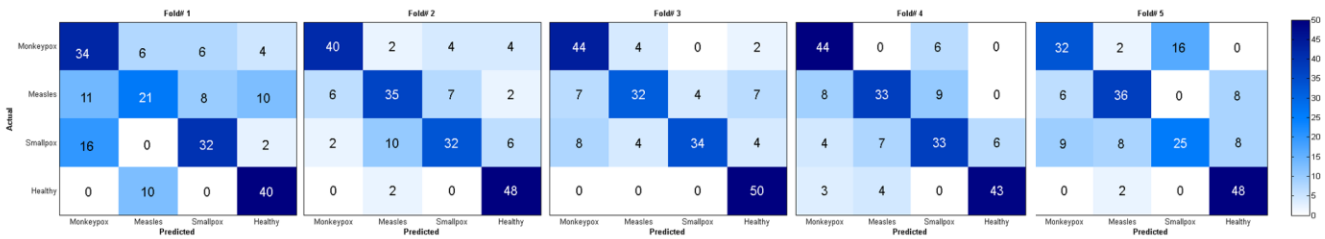
SVM



ResNet50



SqueezeNet



InceptionV3



# Upgraded Very Fast Decision Tree: Energy Conservative Algorithm for Data Stream Classification

Mai Lefa<sup>1</sup>, Hatem Abd-Elkader<sup>2</sup>, Rashed Salem<sup>2</sup>

Department of Information System, Sadat Academy for Management Science, Cairo, Egypt<sup>1</sup>

Department of Information Systems-Faculty of Computers and Information, Minoufia University, Egypt<sup>2</sup>

**Abstract**—Traditional machine learning (ML) techniques model knowledge using static datasets. With the increased use of the Internet in today's digital world, a massive amount of data is generated at an accelerated rate that must be handled. This data must be handled as soon as it arrives because it is continuous, and cannot be kept for a long period of time. Various methods exist for mining data from streams. When developing methods like these, the machine learning community put accuracy and execution time first. Numerous sorts of studies take energy consumption into consideration while evaluating data mining methods. However, this work concentrates on Very Fast Decision Tree, which is the most often used technique in data flow classification, despite the fact that it wastes a huge amount of energy on trivial calculations. The research presents a proposed mechanism for upgrading the algorithm's energy usage and restricts computational resources, without compromising the algorithm's efficiency. The mechanism has two stages: the first is to eliminate a set of bad features that increase computational complexity and waste energy, and the second is to group the good features into a candidate group that will be used instead of using all of the attributes in the next iteration. Experiments were conducted on real-world benchmark and synthetic datasets to compare the proposed method to state-of-the-art algorithms in previous works. The proposed algorithm works considerably better and faster with less energy while maintaining accuracy.

**Keywords**—Classification; energy consumption; Hoeffding bound; Information gain; massive online analysis; stream data; very fast decision tree

## I. INTRODUCTION

In recent years, the amount of generated data is growing significantly. This data is made up of  $n$  ( $n \rightarrow \infty$ ) data samples, and it is described as a series of ordered data sequences, with starting and stopping bytes. Data stream (DS) =  $\{x_1, x_2, \dots, x_i, \dots, x_n\}$ , where  $x_n$  denotes the most recently visible data object in a stream of data [1]. Because the samples arrive in the form of continuous data flow at a great speed, the traditional classifiers cannot access the data instances in real-time [2]. The instances can only be scanned once or stored for a short period of time, so an ideal data stream classifier must be well prepared to deal with a large number of instances in real-time for better classification performance [1,3]. Data stream mining is a subset of machine learning that involves examining data flows that are continuously expanding as time series and developing a classification algorithm based on them. It is highly different from traditional data mining in

terms of processing the mining operation, but it is similar to it in terms of purpose [2].

Machine Learning (ML) techniques consist of three types: unsupervised, semi-supervised, and supervised. Unsupervised algorithms can operate with unlabeled data; the data is clustered into groups with similar characteristics. Semi-supervised algorithms partially work with labeled data [4]. Supervised algorithms require data to be labeled, which they can categorize the data according to a distinct pattern for each class based on the label, such as data stream classification algorithms [5].

Very Fast Decision Tree (VFDT) is an effective classification technique for DS classification, and it has a high level of classification accuracy. It builds the decision tree by learning examples in real-time at a reasonable cost, and its sequential feature can match the data stream's timeliness requirement. To dynamically generate the decision tree, it is based on the enhancement of the Hoeffding tree. The Hoeffding bound ( $\epsilon$ ) is used to make sure that the data utilized to build each sub tree contains enough information [6,7].

The algorithm determines the information gain ( $G$ ) for all noticed attributes after reading the minimum number of examples ( $n_{min}$ ) at that leaf. The two top features are received through using the function *Best Split Suggestion()*. The difference between these features is compared with the  $\epsilon$ , which was calculated previously. As a result, if the difference exceeds the  $\epsilon$ , there will be a split on the tree by changing the leaf with a new internal node. This function is considered among the functions with the highest energy in the algorithm [8,9]. This point is considered the main problem that the paper solves. Therefore, the proposed mechanism focuses on working to reduce bad features that take a lot of energy when calculating their  $G$ , without impacting on the performance of algorithm. These are the motivations and objectives for our mechanism:

- 1) Decrease the amount of calculations by reducing undesirable features on each leaf.
- 2) Maintain the same level of accuracy by deleting just the computations that aren't necessary.

The rest of this paper is structured as follows: Section II clarifies related work. Section III introduces the background. Our algorithm and the corresponding theoretical derivation are explained in Section IV. Experiments and results are covered

in Section V. Finally, the paper is concluded with the conclusions and the future work in Section VI.

## II. RELATED WORK

Many works have gone into improving the VFDT in various ways. The Random Forest algorithm was created by Dong Zhenjiang and Li Lingjuan [10], it is a combinational classifier that performs well in terms of classification. It is made up of multiple decision trees, and it can compensate for the lack of a single decision tree. This technique upgrades the Random Forest algorithm by sliding the time window to match the unlimited data streams, and introduces the criterion of constructing decision trees in the random forest classifier. The limitation of this study is that it improves the accuracy levels without trying to improve other evaluation criteria.

R. J. Lyon et al. proposed the GH-VFDT, a novel classification technique for unbalanced data flows. The Hellinger distance was combined with a stream classifier depending on the Hoeffding bound in an empirical study. They can be combined to create a skew-insensitive decision tree split criterion that enhances minority class recall rates significantly. On unbalanced data, the algorithm can successfully enhance minority class recall rates, with same performance levels [11].

In the research [12], Ariyam Das et al. presented a memory-efficient bootstrap simulation heuristic (Mem-ES) that effectively accelerates the learning process. Experiments show that performing resampling techniques efficiently speeds up node splits for online decision tree learning.

Victor Guilherme Turrisi da Costa et al. proposed two versions of a novel VFDT-based algorithm, SVFDTs were developed to minimize the size of VFDT-induced trees, resulting in a memory-conserving decision tree. Both SVFDTs produce trees that are substantially smaller than those produced by the VFDT, while it doesn't reduce prediction performance statistically [13].

Gayathiri Kathiresan and Krishna Mohanta [14] present the compact method that uses the adaptive reservoir sampling methodology. It helps to reduce memory usage, and handle unbalanced stream data by restricting the information gain deviation based on the improved splitting metric in the information gain measurement. The limitation of this study is that it applied the proposed method to only one dataset, neglecting the other types of datasets with different characteristics, in order to determine the reliability of the findings.

Liang, Chunquan et al. use the Hoeffding bound theory and the Uncertain Naive Bayes classifier (UNB) to improve the speed of construction and classification performance. They use gradual pruning and an adaptive tie-breaking criterion [15]. In a different study, Eva Garcia et al. proposed the *nmin* adaptation approach to enhance parameter adaptation in Hoeffding trees. They dynamically adjust the total amount of examples needed to make a split, this approach saves energy. The limitation of this study is that the *nmin* adaption method adjusts to the best *nmin* parameter based on the assumption, that newly received data would maintain the same distribution as previously seen data [16].

## III. BACKGROUND

### A. Hoeffding Tree

There are several types of supervised classification techniques: k-Nearest Neighbor (kNN), Neural Network, Support Vector Machine (SVM), Naïve Bayes (NB) and Decision Tree (DT). Hoeffding tree algorithm is an extension of the decision tree algorithm for performing DS classification. It reads data in a single pass, and builds model  $b = f(a)$  that maps test example  $a$  to class  $b$ . Every node represents a class feature, and each leaf represents the forecasted class label for that node. Beginning from the root node, the DT grows by exchanging leaf nodes with newly arrived test attributes. It would be the most efficient way of classification [17,18]. VFDT extends the Hoeffding tree algorithm by separating the current best features based on a user-specified threshold value [5].

### B. VFDT

VFDT is a tree-based ML technique for DS based on the Hoeffding bound ( $\epsilon$ ) principles. The nodes represent the features of the dataset. The edges indicate the different possible outcomes for each attribute, and the leaf nodes represent the dataset's class labels. Once the model is complete, test data passes through it, and the decision tree will determine the class label. Each instance is read one by one, it is then sorted into the appropriate leaf, and the statistics are updated [5,19].

At the leaf node, denote  $G(X_i)$  as the heuristic measure of attribute  $X_i$ . The algorithm is used to determine the heuristic measure (information gain ( $G$ )) for all observed features after reading the minimum number of instances (*nmin*) at that leaf. Assume that  $X_a$  and  $X_b$  are the attributes with the best and second-best  $G$  after seeing  $n$  pieces of data. The difference in  $G$  between both the best and second best features ( $\Delta G$ ),  $\Delta G = G(X_a) - G(X_b)$ , is compared to the  $\epsilon$  after calculating it according to Eq. 1. If  $\Delta G > \epsilon$ , then the attribute  $X_a$  is the best attribute of the current leaf node with a probability of  $1 - \delta$ . A node will replace that leaf, and there is a split on the best feature  $X_a$ .

$$\epsilon = \sqrt{\frac{R^2 \ln(\frac{1}{\delta})}{2n}} \quad (1)$$

The feature  $X_a$  is removed from the list of all features  $x$  when calculating the tested attributes in next iterations  $X_m$  as shown in Eq. (2), then the information gain for  $X_m$  must be calculated [20].

$$X_m = x - X_a. \quad (2)$$

The statistics required for attribute splitting are stored in each node. When two discrete attributes have similar split gains  $G$  or the highest and second-highest  $G$  are not significantly different, a tiebreak hyper parameter ( $\tau$ ) is introduced to enable tree growth. This is done by ignoring the Hoeffding bound condition and checking if  $G(X_a) - G(X_b) < \epsilon < \tau$  is true [18]. Because the two top attributes have extremely comparable  $G$  values, the algorithm can split into either of them [21].

The algorithm contains a set of fixed parameters that are determined before the work of the algorithm begins. The  $nmin$  parameter defines the minimum number of samples that the algorithm must observe before computing. If there are sufficient statistics for a good split; the default value is 200. The parameter  $\tau$  is utilized to break a tie in the event of a tie. When the difference between the two features is tiny enough, it suggests that both are equally as good. Therefore, waiting a long time for more instances to make a split is pointless. The  $\delta$  parameter denotes one minus the likelihood of selecting the correct feature to split on. The researchers' default split criterion is information gain or the Gini index [6].

### C. Energy Consumption for VFDT Functions

It is necessary to identify the energy consumption at each function level of the VFDT. The specific functions of the VFDT are four key functions: *homogeneous()*, *label()*, *split()* and *bestSplit()*. If all of the tree's instances can be labeled with a single class, the *homogeneous()* function returns true. The leaf node's label is returned by the *label()* function. The *split()* function encompasses all of the functions involved in splitting an internal node into several children. The best attribute to split on is returned by the *Split()* function. This can be accomplished in a variety of ways, including by utilizing the information gain function [17].

## IV. THEORETICAL MODEL OF UPGRADED U-VFDT

In this paper, U-VFDT is proposed to enhance the VFDT and improve its overall performance, which VFDT is one of the most well-known methods for handling data streams. The splitting is performed based on the current best attributes, whereas the algorithm determines the  $G$  from all noticed attributes after reading the  $nmin$  at that leaf. This process is considered one of the most energy-consuming processes in the algorithm. In this case, the *bestSplit()* function performs unnecessary operations that increase the wasted energy in the algorithm.

The difference in  $G$  between the top and the second-top feature ( $\Delta G$ ) is compared with the Hoeffding Bound ( $\epsilon$ ). If  $\Delta G > \epsilon$ , the leaf is replaced by a node, and there is a split on the best feature. That feature is deleted from the list of features available to split on that branch. This method requires a periodic check because the best attributes are prone to change. The proposed method is introduced to limit the number of checked attributes which are noticed at the leaf. Because the  $G$  of all features is calculated in each split, and it leads to leveling up the algorithm's energy usage. The proposed method introduces a dynamic mechanism for appropriate features group through two basic steps.

*The first step* is adaptive appropriate features.

At this step, the performance of all attributes is analyzed to exclude the attributes with small  $G$ , through calculating the  $G$  for all attributes. If the  $G$  for a specific attribute  $X_i$  is less than the  $G$  of the best attribute  $X_a$  by more than a difference of  $\epsilon$   $G(X_a) - G(X_i) \geq \epsilon$ , this attribute is ignored for that leaf. If  $G(X_a) - G(X_i) \leq \epsilon$ , then the attribute is regarded as top feature.

*The second step* is appropriate features group.

At this step, the appropriate features group ( $P$ ) is used in each leaf node, which stores the top  $X$  features selected at the previous step, and excludes the bad ones. Therefore, the  $P$  surely has captured the true split attribute [19]. This method leads to a reduction of the energy usage, because only appropriate features information gain will be evaluated in each split.

To implement this method, we assumed  $X_a$  is the attribute with highest  $G$ , for any other attribute  $X_i$ .  $X_i$  is said to be appropriate if  $G(X_a) - G(X_i) \leq \epsilon$  and  $X_i$  is shown in the appropriate group  $P$  ( $X_i \in P$ ), other features are removed. After making a split on the best attribute  $X_a$ , the algorithm removes  $X_a$  from the list of features in  $P$  available to split on that branch. In the next iteration, when algorithm calculates  $G$  for all tested attributes in next iterations  $X_m$ , only the list of attributes in the appropriate group will be recalculated, with the deletion of  $X_a$  as shown in Eq. 3.

$$X_m = p - X_a. \quad (3)$$

In theory, this approach should improve accuracy, reduce the number of calculations, and decrease the energy usage. It will be shown during the practical application. In Algorithm 1, a pseudo code displays the implementation of the U-VFDT with appropriate features group.

### Algorithm 1: The U-VFDT with appropriate features group mechanism

#### Require:

- S : the stream of instances
- $\epsilon$  : Hoeffding bound
- $\delta$  : the error probability
- HT: Tree with a single leaf (the root)
- X: set of attributes
- $G(\cdot)$ : split evaluation function
- P : Appropriate features group
- $\tau$ : the tiebreak parameter set by the user

#### Ensure:

- Enhanced Very Fast Decision Tree
- 1. While stream is not empty do
- 2.   Read instance  $I_i$  from S
- 3.   Sort  $I_i$  to corresponding leaf  $l$  using  $\epsilon$
- 4.   Update statistics at leaf  $l$
- 5.   Increment  $n_i$ : instances seen at leaf  $l$
- 6.   If  $n_i \geq nmin$  then
- 7.     Compute  $\epsilon$
- 8.     Compute  $G(X_i)$  for each attribute  $X_i$
- 9.     If  $G(X_a) - G(X_i) \leq \epsilon$  then
- 10.        $X_i \in p$
- 11.       Calculate  $\Delta G(\cdot) \leftarrow G(X_a) - G(X_b)$
- 12.       If  $\Delta G(\cdot) > \epsilon$
- 13.         Split on best attribute  $X_a$  and Replace  $l$  with a node.
- 14.         For each branch of the split do
- 15.           Update new leaves,
- 16.           Add New Leaf  $L_m$  with empty  $p_m$
- 17.           Let  $X_m \leftarrow p - X_a$
- 18.         End for
- 19.       End if
- 20.     End if
- 21.   Else

- |     |                  |
|-----|------------------|
| 22. | Do not split     |
| 23. | Do not update HT |
| 24. | End if           |
| 25. | End while        |

## V. EXPERIMENTS AND RESULTS

In this section, the mechanism is analyzed using reliable datasets that have been utilized in previous researches, with the goal of determining how big of an influence of our algorithm; using the results of its prior efforts as a benchmark, and comparing it with the results of proposed mechanism.

### A. Datasets

The experiment was conducted on several different types of datasets to see the impact of the proposed mechanism on each of those data. There are real-world dataset, synthetic dataset and real-world benchmarks datasets. These datasets were chosen as a baseline for the standard behavior of the algorithm. We obtained the real-world datasets from <https://archive.ics.uci.edu/theml/datasets.php>.

The airlines dataset is the real-world dataset. This classification dataset divides flights into two categories: delayed and not delayed, based on the flight's path and departure and arrival airports. It has eight attributes. The random tree dataset is a synthetic dataset that was created using MOA (Massive Online Analysis).

The synthetic generator generated one million cases. The last two datasets consist of real-world benchmarks, the first one is abalone dataset and this study includes predicting the age of abalone from physical measurements. The second one is adult dataset, which predicts whether income exceeds fifty thousand dollars per year based on census data. Their main characteristics can be seen in Table I.

TABLE I. DATASETS SUMMARY

Datasets	Name	Type	Instances	Numeric features	Binary features
1	airlines	Real-world	539,383	3	5
2	random tree	Artificial	1,000,000	5	5
3	abalone	Real-world benchmark	4177	2	6
4	adult	Real-world benchmark	48842	6	8

### B. Tools

Massive Online Analysis (MOA) is a well-known framework for developing algorithms and conducting experiments. It has a number of ML methods, such as classification, regression, clustering, concept drift detection, and a recommender system. It also includes a number of evaluation tools.

MOA can be employed with WEKA's many classification and clustering approaches. It is used for online real-time stream data, while WEKA is used for offline data [6,22]. The

MOA framework is running in parallel with IPPET (Intel(R) Platform Power Estimation Tool) which can measure how much power different processes are consuming.

A set of specifications for the device utilized in this experiment is also included in the environment for practical application, which impacts energy-related calculations such as the operating system: Windows 7 professional 64-bit (6.1, Build 7601), the processor: Intel(R) core(TM) i3-2350M CPU @ 2.30GHz (4 CPUs), ~ 2.3GHz, and memory: 4096MB RAM.

### C. Experimental Design

The primary goal of the model is to increase the efficiency of VFDT. It focuses on understanding and developing the functions of the algorithm that consume the most energy. An efficient method has been proposed for selecting best attributes which are used to perform splitting. It avoids the calculation of the heuristic measure for unnecessary attributes that consumes high energy.

To the best of our knowledge, there are no previous works that have limited the energy consumption of the VFDT except the reference [16]. It has set  $\epsilon = \Delta G$  in the  $\epsilon$  equation as shown in Eq. (4) to ensure that  $\Delta G \geq \epsilon$  is satisfied during the next iterations, resulting in a split.

$$nmin = \left\lceil \frac{R^2 \ln\left(\frac{1}{\delta}\right)}{2 \cdot \Delta G^2} \right\rceil \quad (4)$$

The following is a comparative study of the practical experiment between the performance of our method U-VFDT, the standard algorithm VFDT, and this previous modification on algorithm P-FVDT. Run-time, power, memory usage, accuracy and energy are measured using four different datasets.

1) *Run-time of VFDT algorithms:* There is a convergence of execution time in our algorithm and other two algorithms as shown in Fig. 1. The increase of examples in random tree dataset leads to an increase in times of the heuristic measure for all attributes; thus, increases the running time. The U-VFDT achieves less time to reach the highest efficiency in random tree dataset. The reason of this result is that U-VFDT works on reducing those attributes that waste time. Also, adult dataset contains a large number of features, so the proposed method is effective in working to reduce these features.

U-VFDT does not achieve effective results in the other two datasets due to two reasons. The first reason is the small airlines dataset features, and the other reason is the few examples of abalone dataset. Thus, our method is not efficient, because it depends on reducing the number of features, where time is wasted. Unlike the VFDT, the time increases dramatically as soon as it receives new examples, whereas heuristic measure is counted for each attribute of a new example when it is received.

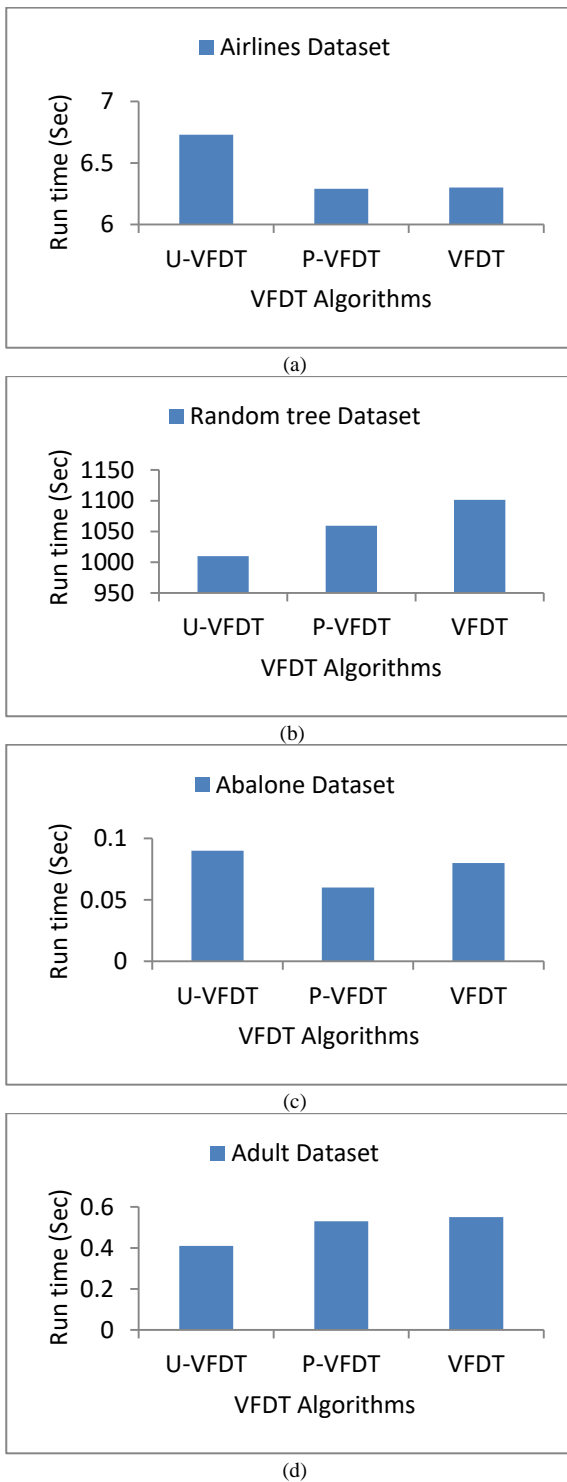


Fig. 1. Run time of VFDT algorithms for all datasets.

2) *Power consumption of VFDT algorithms:* As shown in Fig. 2, lower power values of U-VFDT in some datasets, and other higher values in other datasets. The power levels differ in each algorithm, due to the difference in the nature of the dataset in terms of its examples and features.

U-VFDT does not achieve an effective result in the abalone dataset, because this dataset contains a small number

of examples. U-VFDT proves more efficiency than other two algorithms for other datasets due to two reasons. The first reason is the large number of examples for airlines and random tree datasets, and the other reason is the large number of features for adult dataset. Thus, our method is efficient, because it depends on reducing the number of features.

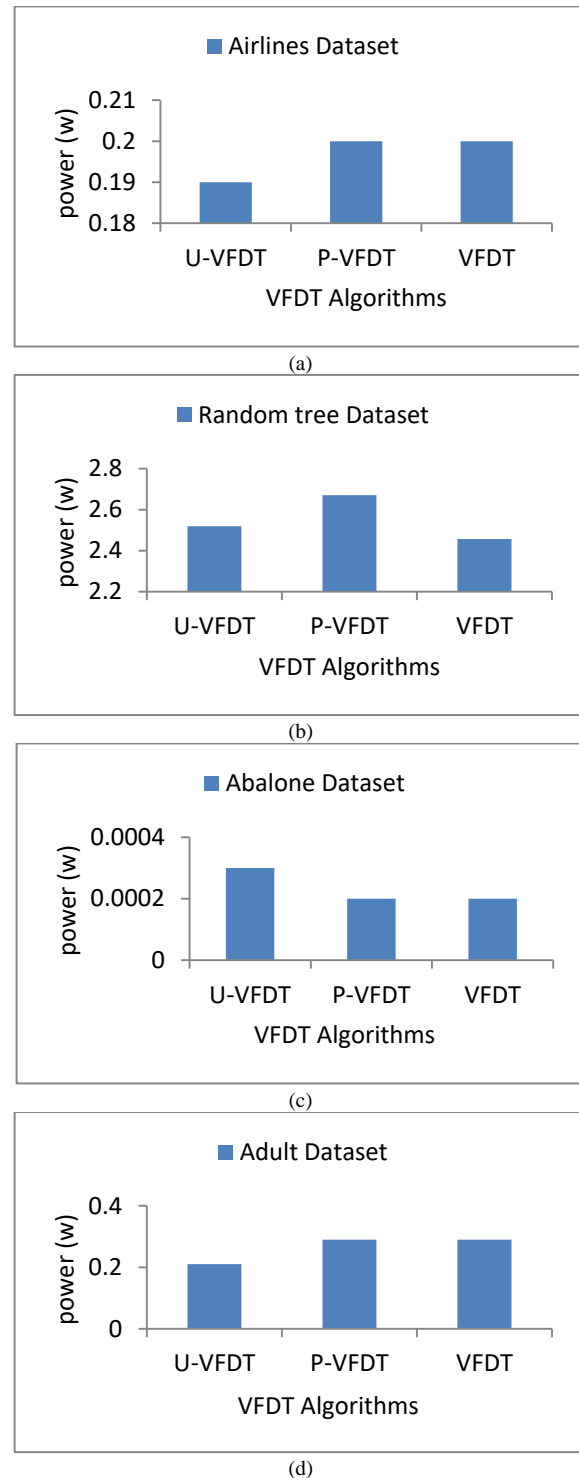
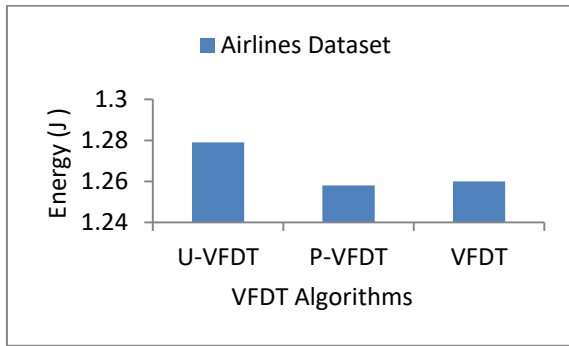


Fig. 2. Power consumption of VFDT algorithms for all datasets.

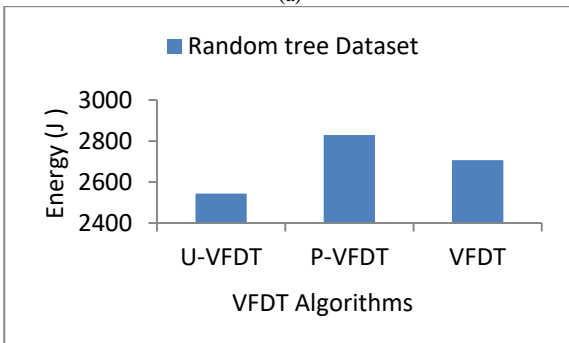
3) Energy consumption of VFDT algorithms for all datasets: U-VFDT succeeded in saving the extra time wasted, which helped in saving energy significantly, based on the Eq. (5). It saves energy only in datasets that contains many examples and attributes, by ignoring the bad features that have no chance of splitting.

$$\text{Energy} = \text{Power} \times \text{Time} \quad (5)$$

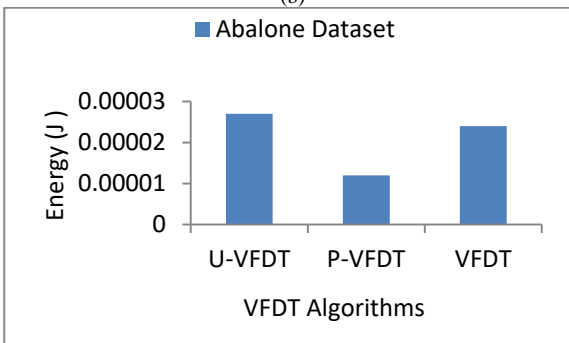
VFDT makes high computations because of the heuristic measure calculations for unnecessary attributes that is required to make a split as shown in Fig. 3.



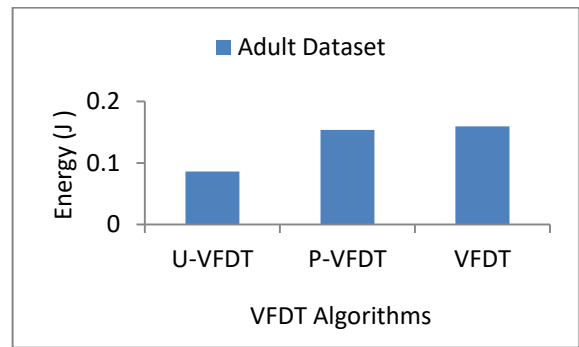
(a)



(b)



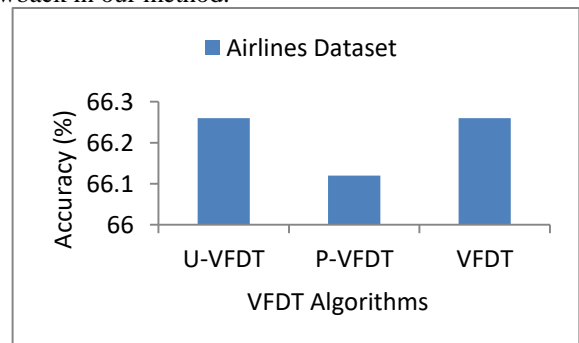
(c)



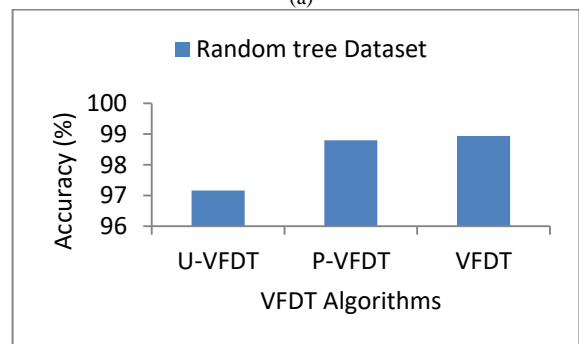
(d)

Fig. 3. Energy consumption of VFDT algorithms for all datasets.

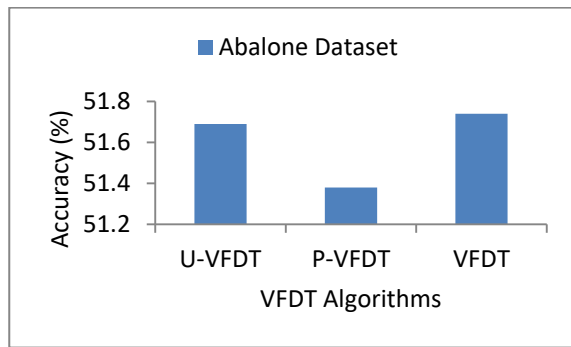
4) Accuracy of VFDT algorithms: As shown in Fig. 4, U-VFDT is maintaining the performance of the algorithm with light impact on accuracy; this is for data with few features and examples such as airlines and abalone datasets. On the contrary, U-VFDT significantly affects the accuracy of the algorithm, but in a greater proportion in the case of data containing a large number of examples and features, such as random tree and adult datasets, which is considered a drawback in our method.



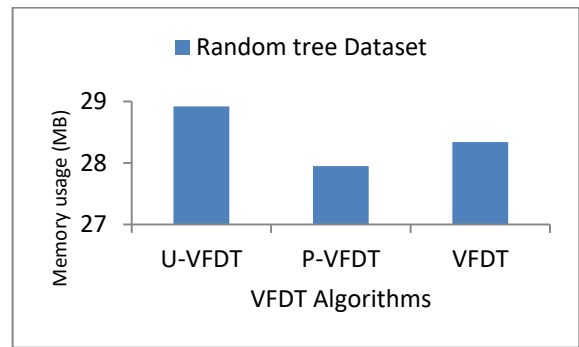
(a)



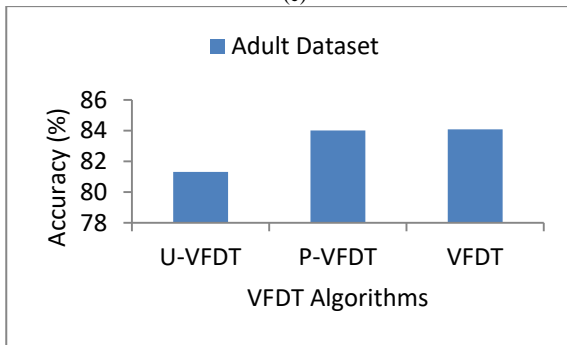
(b)



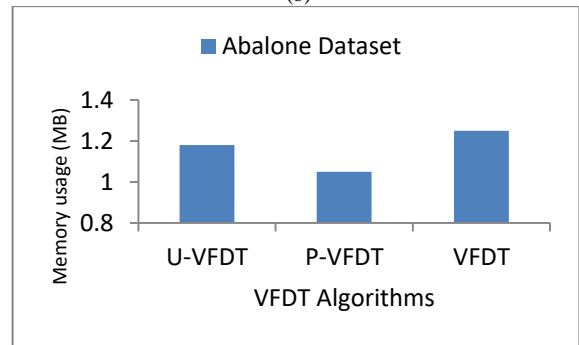
(c)



(b)



(d)

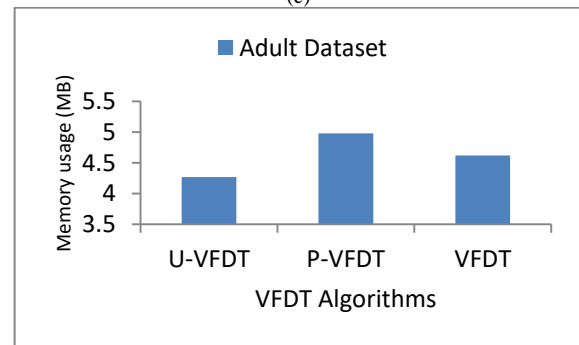


(c)

Fig. 4. Accuracy of VFDT algorithms for all datasets Accuracy (%).

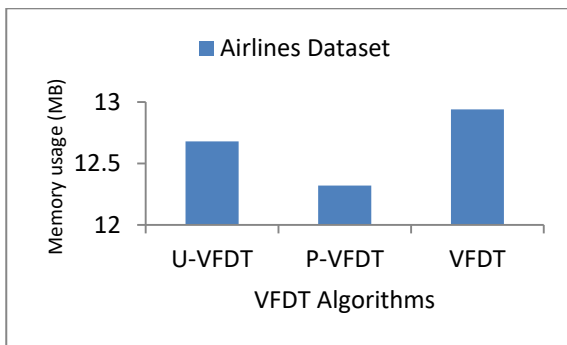
5) *Memory usage of VFDT algorithms:* For memory usage, an adult dataset contains a large number of attributes. U-VFDT proves more efficiency than other two algorithms. It achieves the least required memory for making a split as shown in Fig. 5, which it excludes bad features, thus does not perform any unnecessary operations.

In the random tree dataset, despite the large number of its instances, U-VFDT does not achieve an effective result for it, because this dataset contains noise and concept drift that wastes memory. In the other two small datasets, our method does not achieve an effective result, because it depends on reducing the features.



(d)

Fig. 5. Memory usage of VFDT algorithms for all datasets.



(a)

The experiments clarify that U-VFDT has a better performance than other two algorithms. The proposed algorithm was limited to the heuristic measure of only the good attributes, through which the split could take place. It was excluded of bad attributes that are consuming massive levels of memory, and energy as well as a large run-time. Accordingly, the energy is decreased and the processes are speed up, because of the reduction in both running time and the memory usage. The limitation of this work is that it consumes less energy only in the dataset with a large number of instances and attributes. Also, it does not achieve an effective result for the datasets with noise and concept drift.



The total computational complexity of the VFDT is  $O(n)$ . The computational complexity of the U-VFDT does not exceed its value in the original algorithm, and still retains the same value after modification, since it was  $O(n)$ .

## VI. CONCLUSIONS

This research developed a new technique to improve the VFDT, which allows for an energy conservative algorithm to construct Hoeffding trees without affecting their predictive performance, resulting in lower energy consumption and minor accuracy loss. In VFDT, after the splitting occurred, and the leaf turned into a node, all of features are recalculated to determine which splitting will occur through.

The proposed mechanism recalculates information gain for only the list of attributes in the appropriate group, with the deletion of the feature used for the previous split. It leads to the reduction of unnecessary calculations of bad attributes. Thus, an evolution occurred in the performance of the algorithm in terms of saving time and memory and reducing wasted energy consumption with maintaining the accuracy of the algorithm.

Finally, the mentioned algorithms are compared in different datasets with standard algorithm, and the previous modification work. The U-VFDT used less energy than the VFDT and the P-VFDT only in the datasets with large number instances and attributes. It does not achieve effective results in data that is small in size and has few features. The main work is based on limiting the useless features, thus reducing the number of unnecessary operations that increase running time, energy and memory usage.

Further methods are offered for future work in order to enable an energy-efficient method to build Hoeffding trees for datasets with noise and concept drift without compromising their predictive effectiveness.

## REFERENCES

- [1] Zheng, Xiulin, et al. "A survey on multi-label data stream classification." *IEEE Access* 8 (2019): 1249-1275.
- [2] Li, Xiangjun, et al. "A classification and novel class detection algorithm for concept drift data stream based on the cohesiveness and separation index of Mahalanobis distance." *Journal of Electrical and Computer Engineering* 2020 (2020).
- [3] Rad, Radin Hamidi, and Maryam Amir Haeri. "Hybrid forest: A concept drift aware data stream mining algorithm." *arXiv preprint arXiv:1902.03609* (2019).
- [4] Kok, S., et al. "A comparison of various machine learning algorithms in a distributed denial of service intrusion." *Int. J. Eng. Res. Technol* 12.1 (2019): 1-7.
- [5] Rutuja Jadhav, Neha Sharma "Classification Methods For Data Stream Mining" vol.6. 2018.
- [6] Ashish P. Joshi, Biraj V. Patel." Comparative Study of Different Classification Algorithms for Stream Data Mining Using MOA" *International Journal of Computer Sciences and Engineering*.vol.6. 2018.
- [7] Jia, Shuangying. "A VFDT algorithm optimization and application thereof in data stream classification." *Journal of Physics: Conference Series*. Vol. 1629. No. 1. IOP Publishing, 2020.
- [8] Garcia-Martin, Eva, Niklas Lavesson, and Håkan Grahn. "Identification of energy hotspots: A case study of the very fast decision tree." *International Conference on Green, Pervasive, and Cloud Computing*. Springer, Cham, 2017.
- [9] Garcia-Martin, Eva, Niklas Lavesson, and Håkan Grahn. "Energy efficiency analysis of the very fast decision tree algorithm." *Trends in Social Network Analysis* (2017): 229-252.
- [10] Dong, Z. J., et al. "Random forest based very fast decision tree algorithm for data stream." *Res. Paper* 12 (2017): 52-57.
- [11] Lyon, Robert J., et al. "Hellinger distance trees for imbalanced streams." 2014 22nd International Conference on Pattern Recognition. IEEE, 2014.
- [12] Das, Ariyam, et al. "Learn Smart with Less: Building Better Online Decision Trees with Fewer Training Examples." *IJCAI*. 2019.
- [13] da Costa, Victor Guilherme Turrise, André Carlos Ponce de Leon Ferreira, and Sylvio Barbon Junior. "Strict very fast decision tree: a memory conservative algorithm for data stream mining." *Pattern Recognition Letters* 116 (2018): 22-28.
- [14] Kathiresan, Gayathiri, Krishna Mohanta, and Khanaa VelumailuAsari. "COMPACT: Classifying Stream Data Optimally Using a Modified Pruning and Controlled Tie-threshold." vol.8. 2019.
- [15] Liang, Chunquan, et al. "Learning accurate very fast decision trees from uncertain data streams." *International Journal of Systems Science* 46.16 (2015): 3032-3050.
- [16] García-Martín, Eva, et al. "Hoeffding Trees with nmin adaptation." 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2018.
- [17] Krawczyk, Bartosz, et al. "Ensemble learning for data stream analysis: A survey." *Information Fusion* 37 (2017): 132-156.
- [18] Masrani, Aastha, Madhu Shukla, and Kishan Makadiya. "Empirical Analysis of Classification Algorithms in Data Stream Mining." *International Conference on Innovative Computing and Communications*. Springer, Singapore, 2021.
- [19] Desai, Sharmishta, et al. "Very fast decision tree (VFDT) algorithm on Hadoop." 2016 International Conference on Computing Communication Control and automation (ICCCUBEA). IEEE, 2016.
- [20] Sun, Jiang, et al. "Speeding up very fast decision tree with low computational cost." *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*. 2021.
- [21] Bifet, Albert, et al. "Extremely fast decision tree mining for evolving data streams." *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2017.
- [22] Srimani, P. K., and Malini M. Patil. "Performance analysis of Hoeffding trees in data streams by using massive online analysis framework." *International Journal of Data Mining, Modelling and Management* 7.4 (2015): 293-313.

# Classification Model for Diabetes Mellitus Diagnosis based on K-Means Clustering Algorithm Optimized with Bat Algorithm

Syaiful Anam<sup>1</sup>, Zuraidah Fitriah<sup>2\*</sup>, Noor Hidayat<sup>3</sup>, Mochamad Hakim Akbar Assidiq Maulana<sup>4</sup>  
Mathematics Department, Brawijaya University, Malang, Indonesia<sup>1,2,3</sup>  
Undergraduate Student, Mathematics Department, Brawijaya University, Malang, Indonesia<sup>4</sup>

**Abstract**—Diabetes mellitus is a disease characterized by abnormal glucose homeostasis resulting in an increase in blood sugar. According to data from the International Diabetes Federation (IDF), Indonesia ranks 7th out of 10 countries with the highest number of diabetes mellitus patients in the world. The prevalence of patients with diabetes mellitus in Indonesia reaches 11.3 percent or there are 10.7 million sufferers in 2019. Prevention, risk analysis and early diagnosis of diabetes mellitus are necessary to reduce the impact of diabetes mellitus and its complications. The clustering algorithm is one of methods that can be used to diagnose and analyze the risk of diabetes mellitus. The K-mean Clustering Algorithm is the most commonly used clustering algorithm because it is easy to implement and run, computation time is fast and easy to adapt. However, this method often gets to be stuck at the local optima. The problem of the K-means Clustering Algorithm can be solved by combining the K-means Clustering algorithm with the global optimization algorithm. This algorithm has the ability to find the global optimum from many local optimums, does not require derivatives, is robust, easy to implement. The Bat Algorithm (BA) is one of global optimization methods in swarm intelligence class. BA uses automated enlargement techniques into a solution and it's accompanied by a shift from exploration mode to local intensive exploitation. Based on the background that has been explained, this article proposes the development of a classification model for diagnosing diabetes mellitus based on the K-means clustering algorithm optimized with BA. The experimental results show that the K-means clustering optimized by BA has better performance than K-means clustering in all metrics evaluations, but the computational time of the K-means clustering optimized by BA is higher than K-means clustering.

**Keywords**—Diabetes mellitus; disease diagnosis methods; k-means clustering algorithm; optimization; bat algorithm

## I. INTRODUCTION

Diabetes mellitus is a disease characterized by abnormal glucose homeostasis resulting in an increase in blood sugar. According to data from the International Diabetes Federation (IDF), Indonesia ranks 7th out of 10 countries with the highest number of diabetes mellitus in the world. The prevalence of patients with diabetes mellitus in Indonesia reaches 11.3 percent or there are 10.7 million sufferers in 2019 [1]. Diabetes mellitus causes various complications such as cardiovascular disease, atherosclerotic disease, peripheral neuropathy, diabetic retinopathy, severe foot infections, kidney failure, and sexual dysfunction [2, 3].

Early diagnosis of diabetes mellitus is necessary to reduce the impact of diabetes mellitus and its complications. Clustering algorithms have been used to diagnose and analyze the risk of diabetes mellitus [4-6]. In general, clustering is divided into four categories of use, namely data reduction, hypothesis formation, hypothesis testing, and prediction based on groups [7]. Algorithm clustering is automatically able to recognize patterns in the data so that it can analyze the collected data without the label [8].

The K-mean Clustering Algorithm is the most commonly used clustering algorithm because it is easy to implement and run, the computation time is fast, and easy to adapt [9]. This algorithm has been used in various applications including diagnosis of diabetes mellitus [5], segmentation of diseases in plant leaves [10], heart disease prediction and classification [11, 12] and prediction of diabetes mellitus [13]. However, this method has a drawback, namely random centroid initialization causing, the algorithm to be stuck at the local optima [14]. The clustering result of the K-means algorithm becomes worse because the cluster center is stuck at the local optima. Therefore, the robust initialization of centroid is needed to obtain the good clustering result.

Problems of the K-means Clustering Algorithm can be overcome by combining the K-means Clustering with global optimization algorithms, e.g., swarm intelligence algorithm. This algorithm is able to find the global optimum from many local optimums, does not require derivatives, robust, and easy to implement [15]. Anam et al. have used a swarm intelligence-based algorithm (Particle Swarm Optimization) to segment disease in tomato leaves [16]. One of the swarm intelligence methods is Bat Algorithm, with a faster convergence rate than Genetic Algorithm and Particle Swarm Optimization [17]. This is because BA uses automated enlargement techniques into a promising solution. This enlargement is accompanied by a shift from exploration mode to local intensive exploitation. BA also has been used for many applications, for example travelling salesman problem [18, 19], resource scheduling [20, 21], customer churn [22, 23], brain tumor recognition [24, 25], estimating state of health of lithium-ion batteries [26], detection of myocardial infarction [27] and features selection [28, 29].

\* Corresponding Author

TABLE I. ATTRIBUTES DESCRIPTION OF DATA SET

No.	Attribute Name	Attribute Description
1	HighBP	Respondent has high blood pressure which is decided by health professional.
2	HighChol	Respondent has ever had high blood cholesterol which is decided by health professional.
3	CholCheck	Cholesterol check in the last five years
4	BMI	Body Mass Index (BMI)
5	Smoker	Respondent has smoked at least 100 cigarettes in his/her lifetime.
6	Stroke	Respondent has a stroke.
7	HeartDiseaseorAttack	Respondent who had reported suffering from coronary heart disease or myocardial infarction.
8	PhysActivity	Respondent who reported engaging in physical activity or sports during the last 30 days apart from their regular job.
9	Fruits	Respondent consumes fruit 1 time or more per day
10	Veggies	Respondent consumes vegetable 1 time or more per day
11	HeavyAlcoholConsumption	Heavy drinking or not (adult men drink more than 14 drinks per week and adult women drink more than 7 drinks per week)
12	AnyHealthcare	Possession of any health care coverage, including health insurance, prepaid plans, or government plans.
13	NoDoctorCost	Was there a time in the last 12 months when you needed to see a doctor but couldn't because of costs?
14	GeneralHealth	General health score. [from 1 to 5]
15	MentalHealth	Mental health which includes stress, depression and emotional problems, for how many days during the last 30 days your mental health was not good. [from 1 to 30]
16	PhysicalHealth	Physical health including physical illness and injury, for how many days during the last 30 days your physical health was not good. [from 1 to 30]
17	DiffWalking	Has serious difficulty walking/climbing stairs or not
18	Sex	Indicates the gender of the respondent. [Female: 0, Male: 1]
19	Age	Age category fourteen levels [from 1 to 14]
20	Education	The completed level of education. [from 1 to 6]
21	Income	The household's annual income from all sources: (If the respondent declines at any income level, code "Refuse.")
22	Diabetes	0 is no diabetes, 1 is pre-diabetes or diabetes

Based on the background described, this article proposes the development of a method for diagnosing diabetes mellitus based on the K-means clustering algorithm optimized by BA. The purpose of this research is to develop a rapid method for diagnosing diabetes mellitus using the K-means and BA algorithms. In this article, the K-means algorithm is improved by using the BA algorithm to overcome the problem of the K-means algorithm which is often stuck in the local optima. This research is useful for the prevention and reduction of the impact of diabetes mellitus through a rapid and inexpensive diagnosis of diabetes mellitus by utilizing information technology (machine learning).

## II. PROPOSED METHOD

This sub-chapter will explain the dataset which are used, research stages, stages of the proposed method and evaluation tools used.

### A. Data Set

The dataset used in this study was taken from the web at <https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset>. The dataset was taken from the Behavioral Risk Factor Surveillance System (BRFSS), which is a health-related telephone survey that is collected annually in America.

The dataset consists of several predictor variables, both medical and non-medical, and one target variable (diabetes mellitus sufferers and not diabetes mellitus sufferers). Description of the attributes of the dataset used can be seen in Table I. Class 1 means people with diabetes mellitus or prediabetes while class 0 means not people with diabetes mellitus. This dataset is used for the training process and for evaluation of the built diabetes mellitus prediction model.

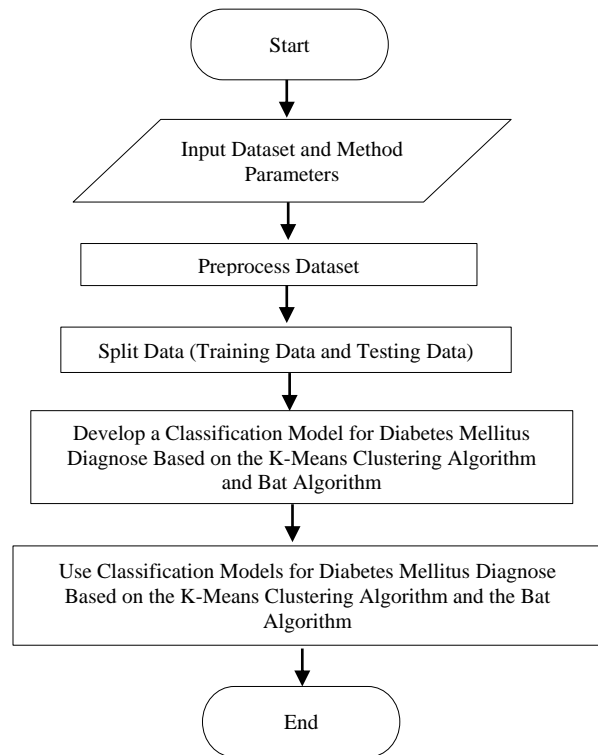


Fig. 1. Flowchart of the proposed method.

### B. Classification Model for Diabetes Mellitus Diagnosis Based on K-Means Clustering Algorithm and Bat Algorithm

This section describes the steps of the Classification Model for Diabetes Mellitus Diagnosis Based on K-Means Clustering Algorithm and Bat Algorithm. The steps or stages of the Classification Model for Diabetes Mellitus Diagnosis Based on the K-Means Clustering Algorithm and the Swarm Intelligence Algorithm can be seen in Fig. 1. The method has several steps which are input dataset and method parameters, preprocess dataset, split data (training data and testing data),

develop a Classification Model for Diabetes Mellitus Diagnosis Based on the K-Means Clustering Algorithm and Bat Algorithm, and use Classification Models for Diabetes Mellitus Diagnosis Based on the K-Means Clustering Algorithm and the Bat Algorithm.

### C. Parameters Setting

Before the Classification Model for Diabetes Mellitus Diagnosis Based on the K-Means Clustering Algorithm and Bat Algorithm is used, there are several parameters that must be set. Some of these parameters include:

- The number of bats used is  $n = 20$ ,
- Maximum number of iterations  $t_{max} = 1000$ ,
- An initial loudness (constant or decreasing)  $A = 1$ ,
- An initial pulse rate (constant or decreasing)  $r_0 = 1$ ,
- Alpha ( $\alpha$ ) = 0.97,
- Gamma ( $\gamma$ ) = 0.1,
- A minimum frequency ( $f_{min}$ ) = 0, and
- A maximum frequency ( $f_{max}$ ) = 2.

These parameters are taken from [17].

### D. Data Preprocessing

Data preprocessing is an initial step in the data mining technique to convert raw data into data that is more efficient and in accordance with the data mining model to be used. Raw data taken from various sources often experience errors, missing values, and are inconsistent, so that the raw data need to be formatted so that data mining results are precise and accurate. In addition, raw data also need to be transformed to change data from its original form into data that is ready to be mined. Data transformation can facilitate the process of extracting data to find new knowledge. One of the data transformation techniques is normalization. Normalization is the process of scaling the attribute values of the data so that they can lie in a certain range. This study uses the Min-Max Normalization Method. The Min-Max Normalization is a normalization method by carrying out a linear transformation of the original data so as to produce a balance of comparison values between the data.

### E. Data Splitting

After preprocessing, dataset is divided into two parts for training and testing. The proportion of training and testing data is 80% and 20%. The training data in this study is used to train the model so as to get a clustering model. Data testing is used to test and evaluate the model, as a simulation of using the model in the real world. Data testing should never be used in model training before to make model validation.

### F. Develop a the Classification Model for Diabetes Mellitus Diagnosis Based on K-Means Clustering Algorithm and Bat Algorithm

The next step is to build the Classification Model for Diabetes Mellitus Diagnosis Based on K-Means Clustering Algorithm and Bat Algorithm. The diagnostic model is built

based on the Clustering Method based on the K-Means Clustering Algorithm and the Bat Algorithm. The first step is to build a K-Means Clustering algorithm that is optimized with the Bat Algorithm. The position of the bat in the Bat Algorithm represents the center of the cluster (centroid). The optimized function in the Bat Algorithm is the objective function of K-means Clustering. After the algorithm is built, the next step is to implement the program with the Python programming language.

The next step is to input the training data, the parameters of the Bat Algorithm and the number of clusters. The training data that will be included in the clustering model training process is the predictor variables of the training data. While the response variable will be used later when evaluating the clustering model after the training phase is complete. Algorithm 1 states the K-means Clustering Algorithm-Based Clustering Method and the Bat Algorithm. Algorithm 2 is used for the association of cluster centers to data classes, while Algorithm 3 is used for the testing process.

### Algorithm 1 Clustering Method Based on K-Means Clustering Algorithm and Bat Algorithm

#### Input:

The training data ( $X_{rain}$ ) with size of  $n \times m$   
The number of cluster ( $K$ )  
The parameters of Bat Algorithm

#### Output:

Best ( $x_*$ ) is the best solution produced

a. Initialize the bat positions and velocities  $x_i^0$  and  $v_i^0$ , ( $i = 1, 2, \dots, N$ ). Each  $x_i$  represents a candidate from the centroid or cluster center. For example, matrix  $C_i$  is the  $i$ -th centroid candidate represented in equation (1).

$$C_i = \begin{bmatrix} c_{1,1} & \dots & c_{1,m} \\ \dots & \dots & \dots \\ c_{K,1} & \dots & c_{K,m} \end{bmatrix} \quad (1)$$

Therefore,  $C_i$  is reshaped to get a matrix of size  $1 \times (m.K)$  and saved in  $x_i^0 = (x_{i,1}, \dots, x_{i,m.K}) = (c_{i,1}, \dots, c_{1,m}, \dots, c_{K,1}, \dots, c_{K,m})$ . It aims to facilitate the calculation process on the Bat Algorithm.  $v_i$  also sized  $1 \times (m.K)$ .

b. Initialize a frequency  $f_i$ , a pulse rate  $r_i$ , and a loudness  $A_i$

c.  $t=0$

d. **while** ( $t < \text{Maximum Iteration}$ ) **do**

1. **for**  $i = 1: N$  **do**

i. Generate the new solutions by adjusting the frequency, updating the velocity and the position of bats using equations (2), (3), and (4).

$$f_i = f_{min} + (f_{max} - f_{min})\beta, \quad (2)$$

$$v_i^{t+1} = v_i^t + (x_i^t - x_*)f_i, \quad (3)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1}. \quad (4)$$

ii. **if** ( $rand < r_i$ ) **then**

Generate the local solutions randomly by

using equation (5),

$$\mathbf{x}_{new} = \mathbf{x}_{old} + \sigma \epsilon_t A^{(t)} \quad (5)$$

where  $\epsilon_t$  is random numbers obtained from the normal Gaussian distribution  $N(0,1)$ ,  $A^{(t)}$  is the average loudness of all bats over time  $t$ , and  $\sigma$  is the scale factor, for simplification, can be used  $\sigma = 0.01$ .

**end if**

- iii. Evaluate the fitness using the objective function of K-means clustering which is stated in equation (6),

$$J = \sum_{j=1}^K \sum_{i=1}^{a_j} \|\mathbf{x}_{train}^j - \mathbf{c}_j\|^2 \quad (6)$$

where  $\mathbf{c}_j$  represents centroid  $j$  of  $K$  centroid.  $\mathbf{c}_j$  reshaping results were obtained  $\mathbf{x}_i$  of size  $1 \times K$ ,  $m$  to matrix  $\mathbf{C}_i$  of size  $K \times m$ .

- iv. **if** ( $rand > A_i$  and  $f(\mathbf{x}_i) < \mathcal{F}(\mathbf{x}_*)$ ) **then**

Update the current solution using one of the solutions from step (i) or (ii)

**end if**

- v. Increase  $r_i$  and reduce  $A_i$  by using Equations (7) and (8),

$$A_i^{t+1} = \alpha A_i^t, \quad (7)$$

$$r_i^{t+1} = r_i^0 [1 - \exp(-\gamma t)], \quad (8)$$

where  $0 < \alpha < 1$ , and  $\gamma > 0$ . According to Yang (2014), to facilitate the search process  $A_i$  and  $r_i$  can equate the value of  $\alpha$  and  $\gamma$ , with value  $\alpha = \gamma = 0.9$ .

- vi. Sort the bats and determine the best solution ( $\mathbf{x}_*$ )

**end for**

**end while**

e. Do a reshape on Best ( $\mathbf{x}_*$ ) to get centroid  $\mathbf{C}$ .

The testing data

The centroids

$y_{testing}$  (class labels of each testing data)

**Output:**

Accuracy, Recall, Precision, *F1 Score*.

1. Calculate the label prediction  $y_{pred}$  based on centroid cluster.
2. Calculate Accuracy, Recall, Precision and *F1 Score*.

### G. Evaluation Metrics

The performance of the proposed method is evaluated by using accuracy, recall, precision and *F1 Score*. The performance of the proposed method is compared to the previous method, namely the K-means Clustering method. If the proposed method is better than the standard method, it can be said that the performance of this method can be improved. The evaluation metrics used are:

- 1) Classification Rate / Accuracy which is calculated using the formulation in equation (9),

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

where TP states that diabetics and is detected as a diabetic. TN stated that healthy person and is detected as a healthy person. FN is the healthy person but detected as diabetics. FP stated that diabetics but detected as a healthy person. Accuracy is used to measure the ratio of correct predictions to the total number of instances evaluated.

- 2) Recall is calculated by the formulation in equation (10). Recall is used to measure the fraction of a correctly classified positive pattern.

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

- 3) Precision is calculated by the formula in equation (11). Precision is used to measure the correctly predicted positive pattern from the total predicted pattern in the positive class.

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

- 4) *F1 Score* is calculated by the formula in equation (12). *F1 Score* is harmonic mean of precision and recall.

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (12)$$

After the evaluation metrics are calculated, the experimental results are analyzed to obtain conclusions.

## III. RESULTS AND DISCUSSIONS

The dataset has different scale on each feature/variable; therefore the algorithm cannot work well. So that, the dataset is needed to be pre-processed to solve this problem by using normalization technique. Furthermore, the data are normalized by using the minmax normalization method. This process will make data that has the same range, namely between values 0 and 1. The normalized data is then divided into two parts, namely, 80% of the data is used for training and the remaining 20% is used for testing. After the data is appropriate with the model to be used, the data can be input for the algorithm to be executed.

### Algorithm 2 Centroid analysis on K-Means Clustering Algorithm-Based Clustering Method and Bat Algorithm

**Input:**

The training data ( $\mathbf{X}_{train}$ ) with size of  $n \times m$

The centroids

$y_{train}$  (class label of each training data)

**Output:**

Accuracy, Recall, Precision, *F1 Score*.

1. Determine which centroid represents the target class based on the majority value of the labels in each cluster.
2. Calculate the label prediction  $y_{pred}$  based on centroid cluster.
3. Calculate Accuracy, Recall, Precision and *F1 Score*.

### Algorithm 3 Testing Algorithm of Classification model of Diabetes Mellitus Diagnosis Based Clustering Method K-Means Clustering and Bat Algorithm

**Input:**

The evaluation tools used to measure the quality of each algorithm are the objective function ( $f_{\min}$ ), accuracy, precision, recall, *F1 score*, and the computational time. The accuracy, precision, recall and *F1 score* are calculated for both training data and testing data. The evaluation tool will reach the optimum value when the objective function is minimum, the accuracy, precision, recall and *F1 score* are maximum, and the computation time is not too long.

The first algorithm to run is the standard K-means Algorithm. The parameters used are the number of clusters of 2 which correspond to the expected number of targets. The iterations are carried out until one of stopping conditions is reached. The parameters used in the Bat Algorithm are initialized with the parameters described in sub-section II.C. This algorithm uses two stopping conditions which are the maximum iteration and convergence condition. The maximum iteration is 1000 times. The algorithm is assumed convergence if the global best doesn't have improvement in 100 iteration. The experiments in this study were repeated 25 times, because the Bat Algorithm and the K-means Algorithm used in the Classification Model for Diabetes Mellitus Diagnosis Based on the K-Means Clustering Algorithm and the Bat Algorithm involve random numbers in obtaining the optimum value of the objective function. Then the average and standard deviation of the evaluation tool used are calculated. The standard deviation is used to measure the spread of recall, accuracy, precision and *F1 scores*, as well as objective function values. While the average value is used to concentrate the results of recall, accuracy, precision and *F1 scores*, as well as objective function values.

Table II shows the comparison of objective function value of Classification Model for Diabetes Mellitus Diagnosis based on the K-means Clustering Algorithm and the Bat Algorithm (training data). It can be shown that the Classification Model for Diabetes Mellitus Diagnosis based on the K-means Clustering Algorithm and the Bat Algorithm has better performance than K-means Clustering Method. It results the smaller the objective values.

TABLE II. COMPARISON OF OBJECTIVE FUNCTION VALUE OF CLASSIFICATION MODEL FOR DIABETES MELLITUS DIAGNOSIS BASED ON THE K-MEANS CLUSTERING ALGORITHM AND THE BAT ALGORITHM (TRAINING DATA)

Method	Average	Deviation Standard
K-means	6389.589	82.41
K-means + Bat Algorithm, $n=10$	6341.573	3.9882
K-means + Bat Algorithm, $n=20$	<b>6340.507</b>	<b>0.0291</b>

TABLE III. MEAN OF EVALUATION METRICS FOR CLASSIFICATION MODEL FOR DIABETES MELLITUS DIAGNOSIS BASED ON THE K-MEANS CLUSTERING ALGORITHM AND THE BAT ALGORITHM (TRAINING DATA)

Method	Accuracy	Precision	Recall	F1 Score	Time (s)
K-means	0.7009	0.70797	0.71500	0.70498	<b>0.02856</b>
K-means + Bat Algorithm, $n=10$	0.72427	<b>0.73632</b>	0.69402	0.71443	1008.645
K-means + Bat Algorithm, $n=20$	<b>0.72431</b>	0.73459	<b>0.69744</b>	<b>0.71553</b>	2794.129

TABLE IV. DEVIATION STANDARD OF EVALUATION METRICS FOR CLASSIFICATION MODEL FOR DIABETES MELLITUS DIAGNOSIS BASED ON THE K-MEANS CLUSTERING ALGORITHM AND THE BAT ALGORITHM (TRAINING DATA)

Method	Accuracy	Precision	Recall	F1 Score	Time (s)
K-means	0.06834	0.07391	0.09370	0.04588	0.00927
K-means + Bat Algorithm, $n=10$	0.002607	0.003948	0.014368	0.006265	544.6731
K-means + Bat Algorithm, $n=20$	0.000591	0.00041	0.002546	0.001166	4542.019

TABLE V. MEANS OF EVALUATION METRICS FOR CLASSIFICATION MODEL FOR DIABETES MELLITUS DIAGNOSIS BASED ON THE K-MEANS CLUSTERING ALGORITHM AND THE BAT ALGORITHM (TESTING DATA)

Method	Accuracy	Precision	Recall	F1 Score
K-means	0.6956	0.7010	0.6550	0.6764
K-means + Bat Algorithm, $n=10$	0.7155	<b>0.7311</b>	0.6938	<b>0.7118</b>
K-means + Bat Algorithm, $n=20$	<b>0.7156</b>	0.7128	<b>0.6971</b>	0.70471

TABLE VI. DEVIATION STANDARD OF EVALUATION METRICS FOR CLASSIFICATION MODEL FOR DIABETES MELLITUS DIAGNOSIS BASED ON THE K-MEANS CLUSTERING ALGORITHM AND THE BAT ALGORITHM (TESTING DATA)

Method	Accuracy	Recall	Precision	F1 Score
K-means	0.06068	0.06819	0.04169	0.04988
K-means + Bat Algorithm, $n=10$	0.00146	0.02574	0.01117	0.01363
K-means + Bat Algorithm, $n=20$	0.00083	0.02778	0.00276	0.0136

Table III shows the mean of evaluation metrics for Classification Model for Diabetes Mellitus Diagnosis based on the K-means Clustering Algorithm and the Bat Algorithm (training data), while Table IV shows the deviation standard of evaluation metrics for Classification Model for Diabetes Mellitus Diagnosis based on the K-means Clustering Algorithm and the Bat Algorithm (training data). The experimental results show that the that the Model for Diabetes Mellitus Diagnosis based on the K-means Clustering Algorithm and the Bat Algorithm are superior to the standard K-means method. that the Model for Diabetes Mellitus Diagnosis based on the K-means Clustering Algorithm and the Bat Algorithm are superior in all evaluation tools (accuracy, recall, precision and *F1 Score*). The standard deviation of the accuracy, recall, precision and *F1 score* of the Diabetes Mellitus Diagnostic Model Based on the K-Means Clustering Algorithm and the Swarm Intelligence Algorithm is very small. This means that this method results in low variation. The computational time required for training the Model for Diabetes Mellitus Diagnosis based on the K-means Clustering Algorithm and the Bat Algorithm is much longer than the standard K-means method.

Table V shows the results of the classification model evaluation for data testing. The experimental results show that the Classification Model for Diabetes Mellitus Diagnosis based on the K-means Clustering Algorithm and the Bat Algorithm yields much better results compared to the standard K-means method for accuracy, recall, precision and *F1 Score values*. Tables V and VI also show the accuracy, recall, precision and *F1 scores* produced by the Model for Diabetes

Mellitus Diagnosis based on the K-means Clustering Algorithm and the Bat Algorithm are not much different between training and testing data. This shows that the proposed method has good performance and neither overfitting nor underfitting occurs.

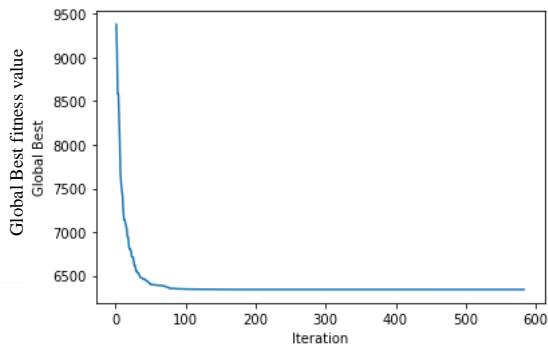


Fig. 2. Graph of iteration and global best relationships of the diabetes mellitus diagnostic method based on the K-Means clustering algorithm and the swarm intelligence algorithm with 20 particles (global best fitness value 6340.499).

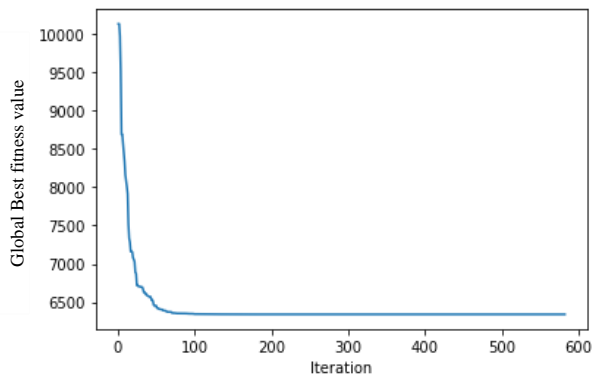


Fig. 3. Graph of iteration and global best relationships of the diabetes mellitus diagnostic method based on the K-Means clustering algorithm and the swarm intelligence algorithm with 10 particles (global best fitness value = 6340.497).

Fig. 2 and 3 show graphs of the iteration and global best relationships of the Classification Model for Diabetes Mellitus Diagnosis based on the K-means Clustering Algorithm and the Bat Algorithm with 20, and 10 Particles. The figures show the method converges less than 600 iterations for the number of particles 20 and 10. The global best convergent fitness value is around 6340.5.

#### IV. CONCLUSIONS

Based on the experimental results and analysis of the experimental results, several conclusions were obtained. that the Model for Diabetes Mellitus Diagnosis based on the K-means Clustering Algorithm and the Bat Algorithm developed from the K-Means Clustering Algorithm by adding the Bat Algorithm optimizer to determine the centroid of the cluster. The experimental results revealed that the number of bats has an effect on the method's convergence speed and processing time. The experimental results reveal that the Model for Diabetes Mellitus Diagnosis based on the K-means Clustering Algorithm and the Bat Algorithm are able to diagnose diabetes

mellitus quite well. The accuracy obtained is around 72.4% and the *F1 score* is 71.4% for training data, and the accuracy obtained is around 71.55% and the *F1 score* is 71.18% for data testing. The evaluation results show that the performance of the Model for Diabetes Mellitus Diagnosis based on the K-means Clustering Algorithm and the Bat Algorithm is better than the standard K-means for evaluating accuracy, precision, recovery, f1 score, but the recovery time is quite large.

#### ACKNOWLEDGMENT

We would like to thank Brawijaya University for funding this research through the Hibah Penelitian Pemula (HPP) research grant.

#### REFERENCES

- [1] InfoDatin, "Pusat data dan informasi Kementerian Kesehatan RI", 2020.
- [2] K. Papatheodorou, M. Banach, M. Edmonds, N. Papanas, and D. Papazoglou, "Complications of diabetes", *Journal of Diabetes Research*, vol. 2015, 2015, <http://dx.doi.org/10.1155/2015/189525>
- [3] D. Tomic, J. E. Shaw, and D. J. Magliano, "The burden and risks of emerging complications of diabetes mellitus", *Nature Reviews Endocrinology*, vol. 18, 2022, pp. 525–539.
- [4] H. Wu, S. Yang, Z. Huang, J. He, and X. Wang, "Type 2 diabetes mellitus prediction model based on data mining", *Informatics in Medicine Unlocked*, vol. 10, 2018, pp. 100–107
- [5] T. Santhanam and M. S. Padmavathi, "Application of k-means and genetic algorithms for dimension reduction by integrating SVM for diabetes diagnosis", *Procedia Computer Science*, vol. 47, 2015, pp. 76 – 83
- [6] A. M. Simarmata, M. A. P. Sianipar, S. Singh, I. I. M. Gulo, and J. B. R. Purba, "Grouping diabetes diagnosis based on age range with k-means algorithm", *Jurnal Mantik*, vol. 5, no. 2, 2021, pp. 1408–1412
- [7] S. Theodoridis, and K. Koutroumbas, "Clustering: basic concepts", *Pattern Recognition*, Edisi 4, Academic Press, 2009, pp. 595–625. doi:10.1016/b978-1-59749-272-0.50013-x
- [8] Y. Lei, Z. He, Y. Zi, and X. Chen, "New clustering algorithm-based fault diagnosis using compensation distance evaluation technique", *Mechanical Systems and Signal Processing*, vol. 22, 2008, pp. 419–435.
- [9] M. Mancas, and B. Gosselin, "Fuzzy tumor segmentation based on iterative watersheds", *Proc. STW Conf. of ProRISC*, Veldhoven, Netherlands, 2003
- [10] S. Zhang, H. Wang, W. Huang, and Z. You, "Plant diseased leaf segmentation and recognition by fusion of superpixel, k-means and PHOG", *Optik*, vol. 157, 2018, pp. 866–872
- [11] M. Thangamani, R. Vijayalakshmi, M. Ganthimathi, M. Ranjitha, P. Malarkodi and S. Nallusamy, "Efficient classification of heart disease using k-means clustering algorithm", *International Journal of Engineering Trends and Technology*, vol. 68, no. 12, 2020, pp. 48–53.
- [12] R. Shinde, S. Arjun, P. Patil and J. Waghmare, "An intelligent heart disease prediction system using K-means Clustering and Naive Bayes Algorithm", *International Journal of Computer Science and Information Technologies*, vol. 6, no. 1, 2015, pp. 637–639.
- [13] C. Fiarni, E. M. Sipayung and S. Maemunah, "Analysis and prediction of diabetes complication disease using data mining algorithm", *Procedia Computer Science*, vol. 161, 2019, pp. 449–457
- [14] K., Singh, D. Malik, and N. Sharma, "Evolving limitations in k-means algorithm in data mining and their removal," *International Journal of Computational Engineering & Management*, vol. 12, 2011, pp. 105–109.
- [15] A. K Kordon, "Swarm intelligence: the benefits of swarms", *Applying Computational Intelligence*, 2009, pp. 145–174.
- [16] S. Anam, and Z. Fitriah, "Early blight disease segmentation on tomato plant using k-means algorithm with swarm intelligence-based algorithm", *International Journal of Mathematics and Computer Science*, vol. 16, no. 4, 2021

- [17] X. Yang, "A new metaheuristic bat-inspired algorithm", *Nature Inspired Cooperative Strategies for Optimization*, Springer, 2010, pp 65-74.
- [18] E. Osaba, X. Yang, F. Diaz, P. Lopez-Garcia, and R. Carballedo, "An improved discrete bat algorithm for symmetric and asymmetric Traveling Salesman Problems", *Engineering Applications of Artificial Intelligence*, vol. 48, 2016, pp. 59-71.
- [19] Y. Saji and M. Barkatou, "A discrete bat algorithm based on Lévy flights for Euclidean traveling salesman problem", *Expert Systems with Applications*, vol. 172, 2021, 114639.
- [20] L. Jacob, "Bat Algorithm for resource scheduling in cloud computing", *International Journal for Research in Applied Science and Engineering Technology*, vol. 2 (4), 2014, pp. 53-57.
- [21] J. Zheng and Y. Wang, "A hybrid multi-objective bat algorithm for solving cloud computing resource scheduling problems", *Sustainability*, vol. 13, pp. 1-25.
- [22] S. Induja, and V. P. Eswaramurthy, "Customer churn prediction and attribute selection in telecom industry using kernelized extreme learning machine and bat algorithms", *International Journal of Science and Research*, vol. 5 (12), 2016, pp. 258-565.
- [23] M. Li, | C. Yan, W. Liu and X. Liu, "An early warning model for customer churn prediction in telecommunication sector based on improved bat algorithm to optimize ELM", *International Journal of Intelligent Systems*, vol. 36, no. 7, 2021, pp. 1-28.
- [24] R. Chawla, S. M. Beram, C. R. Murthy, T. Thiruvankadam, N. P. G. Bhavani, R. Saravanakumar, and P. J. Sathishkumar, "Brain tumor recognition using an integrated bat algorithm with a convolutional neural network approach", *Measurement: Sensors*, vol. 24, 2022.
- [25] G. R. Sreekanth, A. F. Alrasheedi, K. Venkatachalam, M. Abouhawwash, and S. S. Askar, "Extreme Learning Bat Algorithm in brain tumor classification", *Intelligent Automation & Soft Computing*, vol. 34, no.1, 2022, pp. 249-265.
- [26] D. Ge, Z. Zhang, X. Kong and Z. Wan, "Extreme Learning Machine using Bat Optimization Algorithm for estimating state of health of lithium-ion batteries", *Appl. Sci.* 2022, 12, 1398
- [27] P. Kora and S. R. Kalva, "Improved Bat algorithm for the detection of myocardial infarction", *Springer Plus*, vol. 4, 2015, pp. 1-18.
- [28] S. Jeyasingh and M. Veluchamy, "Modified Bat Algorithm for feature selection with the Wisconsin Diagnosis Breast Cancer (WDBC) dataset", *Asian Pacific Journal of Cancer Prevention*, vol. 18, no. 5, pp. 1257-1264.
- [29] B. Yang, Y. Lu, K. Zhu, Gu. Yang, J. Liu and H. Yin, "Feature Selection Based on Modified Bat Algorithm", *IEICE Trans. Inf. & Syst.*, vol. E100-D, no. 8, 2017, pp. 1860-1869.



# Descriptive Analytics and Interactive Visualizations for Performance Monitoring of Extension Services Programs, Projects, and Activities

Noelyn M. De Jesus<sup>1</sup>, Lorissa Joana E. Buenas<sup>2</sup>

Assistant Professor, College of Informatics and Computing Sciences<sup>1</sup>

Professor, College of Informatics and Computing Sciences<sup>2</sup>

Batangas State University ARASOF-Nasugbu, Nasugbu, Batangas, Philippines<sup>1,2</sup>

**Abstract**—Providing universities with high technology-enabled automation tools to support the administrative decision-making processes will enable them to achieve their objectives. For an institution to succeed in its everyday tasks, it should come up with the emerging and modernized management services constituted, among others, by cloud, mobile, and business analytics technology. With this, the institution's operations and management efficiency are ensured. This study aims to develop a system with descriptive analytics named as MET Online Services that will automate and optimize the monitoring of extension services key performance indicators (KPIs) in order to help the institution in making better, data-driven decisions. The dashboards and interactive visualizations of the developed system will provide quick access to real-time progress of the extension services programs, projects, and activities. Results interpretation depicted that the developed system is indeed feasible for implementation, proven to be fully-functional and passed the quality software standards of a Certified Software Quality Assurance Specialist. As the developed system satisfied the users' expectations and requirements, it would be an effective tool for the institution, extension services unit, and community, to make better strategic decisions and continuously deliver quality services to the community.

**Keywords**—Business analytics; descriptive analytics; dashboards; interactive visualizations; extension services; monitoring; key performance indicators; KPIs; community

## I. INTRODUCTION

Managing an institution and bringing together departments and campuses to achieve the vision and quality objectives are always a big challenge for the university management. Providing universities with high technology-enabled automation tools to support the administrative decision-making processes enable them to achieve their objectives. And for an institution to succeed in its everyday tasks, it should cope up with the emerging and modernized management services [1] constituted, among others, by cloud, mobile, and business analytics technology. It is supported by [2, 3], that Higher education institutions are expanding ICT-driven projects in both developed and developing nations to improve quality services in education and increase efficiency and effectiveness. With this, the institutions' operations and management efficiency are ensured.

An effective management system can help an organization use its resources more efficiently and improve its financial performance, risk management, protection of people and the environment, and ability to provide consistent and better services and products. This makes the organization more valuable to its customers and other stakeholders [4]. And to facilitate better flow of information, management system with the use of data analytics within an organization can aid in decision-making that is more strategically oriented, more effective, and maximizes performance [5, 6].

According to [7], management system is the process through which a company oversees the interdependent elements of its operations in order to achieve its goals. These goals may have to do with a variety of things, such as the continuous improvement of the products or services, operational effectiveness, performance, health and safety at work. In the Philippines, State, Universities, and Colleges are expected to extend their academic and community services to the community in accordance with the Commission on Higher Education's (CHED) directives along with instruction, research and production [8]. The office's main goal is to establish a sustainable partnership with public and private organizations with the purpose of reducing poverty. As [9] explicates that there are concerns about implementing extension programs where academic institutions must closely monitor and assess the results of their grassroots community engagement in order to attain development goals.

In this study, a monitoring system with descriptive analytics was developed and will serve as a tool in providing the users with beneficial information to do the extension activities effectively and efficiently that will allow the extension services unit team plan, implement, analyze (at both tactical and strategic levels) and administer all extension services procedure as a mechanized system in the operational level. It will automate and optimize the monitoring of extension services key performance indicators (KPIs) in order to help the institution in making better, data-driven decisions through dashboards, interactive visualizations, and reports. The study also aims to pass in the technical evaluation of a Certified Software Quality Assurance Specialist and determine the level of satisfaction and acceptance of the end-users towards the developed system.

## II. RELATED WORK

### A. Extension Services

According to [10], extension is a relatively new phenomenon and is defined as the intentional transfer of skills and the organized exchange of information. Extension service is a community-based educational opportunity offered by academic institutions [11], with the primary objective of establishing long-term partnerships with both public and private organizations working toward poverty alleviation. The office is in charge of carrying out the university's program for outreach and community development. It makes it simpler to put into action policies and programs that are intended to empower communities, especially those who have been underserved. As [12] stated that the educators from various universities came together and developed community and extension programs in the university. Serving the community is one of an organization's key principles.

The Philippine Higher Education system's objectives include improving the quality of life for Filipinos, adapting effectively to changing societal requirements and circumstances, and offering answers to issues at the local, regional, and national levels, according to [13]. As a core responsibility of universities and colleges, Extension services have been measured or evaluated in several state, universities, and colleges' assessments, including SUC Levelling, the Annual Major Final Outputs of SUCs, CHED's Institutional Sustainability Assessment, and the Accrediting Agency of Chartered Universities and Colleges (AACUP) Institutional and Program Accreditation [14].

According to [15], extension services unit should encourage, motivate, and elevate beneficiary awareness, participation, and satisfaction, especially for low-rated program services. By developing and improving its extension services to be more responsive to the community, the Philippine Higher Education Institutions (HEIs) participated in the discussion of the 2017-2022 Philippine Development Agenda (PDA). [16] explains that the government will create a strong national environment for innovation and knowledge creation, which is necessary for the nation's engagement in the global knowledge economy. It will make investments in: (a) improving the research capabilities of college and university faculty, research staff, and graduate students; (b) building the capital and institutional infrastructure needed for knowledge production and innovation; and (c) building up, retraining, and retaining a sustainable stream of new researchers.

Similar to this, the government will support formal collaborations between HEIs and community, commercial, and industry stakeholders. The intention is to integrate "informal" community-based learning and innovation with "formal" academic research and innovation as a means of accomplishing this goal. Policies and procedures will be developed for the operationalization of extension as part of applied research engagement as well as for the normative and appropriate distribution of hours for teaching and research.

### B. Key Performance Indicators (KPIs)

A performance indicator, also known as key performance indicator (KPI) is an approach of measuring performance [17].

It examines the extent to which an organization was successful in carrying out a program, a project, a product, or some other initiative in which it was involved [18].

There have been many companies that have used inappropriate measurements; as a result, metrics of this kind should not be labeled performance indicators. According to [19], the majority of businesses that are regarded as having a genuine control system make use of key performance indicators. This is the reason why there are accountants, company executives, and consultants who have appropriate knowledge and experience regarding Key Performance Indicators.

One simply cannot place an adequate amount of emphasis on the significance of key performance indicators (KPIs) to the overall success of an organization. Key performance indicators (KPIs) are metrics that evaluate how well a firm is currently doing in relation to its overall goals [20].

Key performance indicators focus on the improvement of strategic and operational levels, which provide an analytical framework for decision-making, and raise knowledge and interest in many areas [21]. KPI plays an important role because it receives timely and accurate information by comparing current performance to a goal that is necessary to satisfy business needs and goals. This allows KPI to play an instrumental role in ensuring that business needs and goals are met. Alignment with corporate goals, active monitoring, report accuracy, and timeliness are all recommended KPI selection techniques [22]. KPI reporting and dashboards are automated to keep all stakeholders on the same page.

The process of managing utilizing key performance indicators also includes the steps of establishing goals (the expected level of performance) and evaluating progress toward those goals. According to [23], to have good KPIs the organization should perform the following, (a) provide objective evidence of progress towards achieving a desired result; (b) measure what is intended to be measured to help inform better decision making; (c) offer a comparison that gauges the degree of performance change over time, (d) track efficiency, effectiveness, quality, timeliness, governance, compliance, behaviors, economics, project performance, personnel performance or resource utilization; and (e) balance the leading and lagging indicators.

Key performance indicators serve to specify a predetermined set of values that have been established. The author in [24] refers to these unprocessed groups of numbers as indicators since they can be used to feed data aggregation systems. Measuring key performance indicators (KPI) can take either a quantitative or qualitative approach. A quantitative measurement compares facts to a standard and uses facts that have a specific objective numerical value. While the qualitative, denoting conformity to a standard that is not quantifiable or an interpretation of one's own feelings, preferences, beliefs, or experiences

### C. Descriptive Analytics

Business analytics finds patterns and trends in massive data sets to improve decision-making and performance. In corporate reports and dashboards, descriptive analytics is used to produce

key performance indicators (KPIs) and metrics. Among the various sorts of analytics, descriptive analytics was recognized as the one that was the most fundamental and the one that was utilized the most [25]. Reports, key performance indicators (KPIs), and other business measures that allow businesses to monitor performance and other trends can be generated using descriptive analytics.

Four categories are frequently used to classify analytics technologies [26]: (i) descriptive (what happened), (ii) diagnostic (why did it happen), (iii) predictive (what will happen next), and (iv) prescriptive (what should be done about it). Academic institutions usually place a strong emphasis on the potential benefits of technology that makes use of sophisticated modeling in order to be either prescriptive or predictive [27, 28, 29, 30]. The majority of adopting companies' sales and marketing teams frequently employ descriptive analytics and straightforward dashboards for key performance indicators (KPIs) [31, 32, 33].

Descriptive analytics are utilized in numerous facets of a business by companies in order to assess the quality of their operations and determine whether or not they are on track to achieve their organizational objectives and used in various ways, particularly in reports, visualizations, and dashboards.

#### D. Reports

Management Information System Report is an umbrella phrase that describes a group of reports that give a picture of what goes on in an organization on a daily basis, which enables the functions of your firm to be studied. MIS Report is an acronym that stands for Management Information System [Futrlı]. They are a great tool for evaluating the performance of a firm and making decisions that are based on accurate and up-to-date information. They comprise of numerous separate reports that cover various areas of a business.

#### E. Visualizations

According to [34], The visualization of data through the use of charts, graphs, and maps, which can indicate trends in data as well as dips and spikes in a fashion that is obvious and easily accessible to the audience, is an excellent choice for providing a suitable outlet for presenting descriptive analysis. It is supported by [35] make information easier to perceive and understand. Data visualization involves the process of incorporating facts or information into a pictorial or graphical environment, like charts, maps, or other visual formats.

#### F. Dashboards

On a consistent basis, dashboards provide a quick overview of key performance indicators (KPIs) that are pertinent to a certain objective or organizational system (e.g., income, marketing, human resources, or production). In real-world usage, the terms "progress report" and "report" can sometimes be referred to as "dashboard" [36].

Regularly, the "dashboard" is shown on an internet page that is connected to a database. This connection makes it

possible for the document to be continuously updated. The term "dashboard" originated from the instrument panel seen in automobiles, where the driver can see, at a glance, the most critical functions displayed on the instrument cluster [37].

Managers can monitor the contributions of varied departments using digital dashboards. Digital dashboards can catch and record distinct information points from every department to provide a "snapshot" of overall performance. Benefits of using digital dashboards include visual presentation of overall performance measures, ability to identify and correct negative tendencies, degree efficiencies/inefficiencies, ability to generate precise reports showing new traits, ability to make more informed decisions based on accumulated business intelligence, align techniques and organizational goals, save time compared to walking a couple of reviews, and benefit total visibility of all measures.

#### G. Classification of Dashboards

Dashboards can be broken down according to position, and they combine strategic analysis with informational, analytical, and operational perspectives according to [38]. Strategic dashboards are a useful tool for managers at any level in an organization, and they give the concise evaluation that decision makers require in order to monitor the health of the business as well as its potential opportunities. These kinds of dashboards focus on high-level measurements of overall performance as well as forecasts. Static snapshots of data (daily, weekly, monthly, and quarterly) that are not constantly changing from one moment to the next are an advantage for strategic dashboards [39].

Dashboards for analytical purposes frequently include greater context, comparisons, and records, along with subtler overall performance evaluators. Analytical dashboards usually assist interactions with the data, along with drilling down into the underlying information. When dashboards were first introduced, its primary function was monitoring; but, as technology has advanced, their use has expanded to include more analytical tasks. Dashboards are currently being used in a way that incorporates scenario analysis, drill down capabilities, and flexible presentation format options [40].

### III. METHODOLOGY

This section highlights the technical system architecture and explains the composition of the developed system as well as the roles involved for the utilization of the system functionality.

#### A. MET Online Services Technical System Architecture

Fig. 1 shows the technical system architecture of the developed system, which consists of the roles of the actors, business services, business processes, application services, technology services, and technology components.

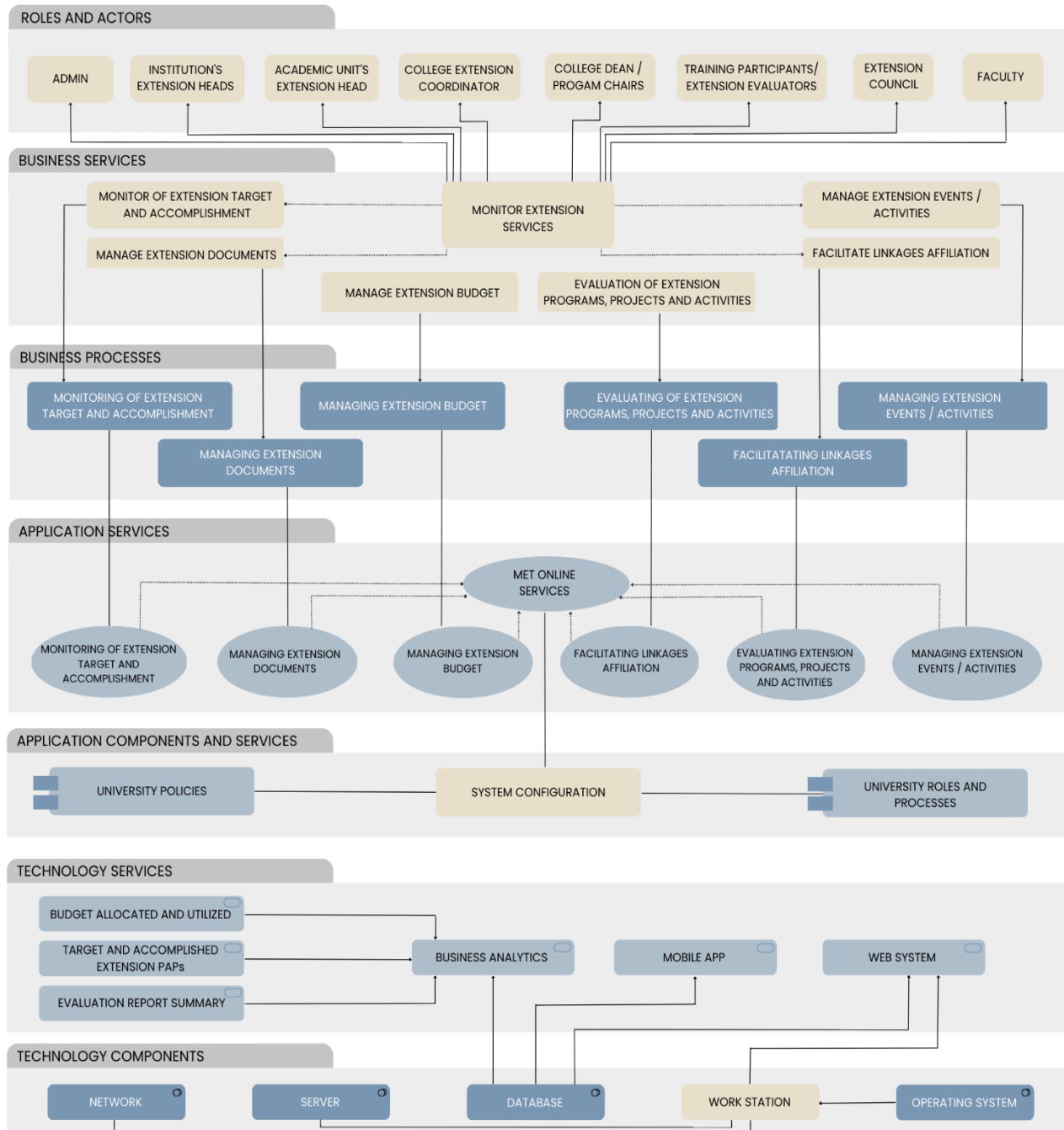


Fig. 1. Technical system architecture

### B. Roles and Actors

In order to make the most of and maximize the usability of the system, it displays its contents and components that are employed. MET Online Services is composed of eight (8) major roles and actors to wit: System Administrator, Institution's Extension Head, Academic Unit's Extension Head, College Extension Coordinators, College Dean / Program Chairpersons, Training Participants, Extension Council, and Faculty. The following users have a corresponding role in order to utilize the system.

### C. Business and Application Services

The business services / processes; and application services of MET Online Services has six (6) major components embedded in the extension services unit, these are as follows: (a) monitoring of extension programs, projects, and activities (PPAs) target and accomplishment, (b) managing extension services documents, (c) managing extension budget, (d) evaluating the extension programs, projects, and activities, (e) facilitating linkages affiliation, and (f) managing extension events / activities.

#### D. Technology Services and Components

For the technology services, the developed system is supported by business analytics, mobile application and web service. The business analytics is comprised of target and accomplished extension PPAs, budget allocated and utilized and evaluation report summary, where in technology services of the developed system it is the main feature along with the mobile application and web service.

In order to utilize the functionality of the developed system, the management must provide a secured network, normalized database, and an operating system that is suitable for the system.

### IV. RESULTS AND DISCUSSIONS

This study developed a system with descriptive analytics named as “MET Online Services” that will automate and optimize the monitoring of extension services key performance indicators (KPIs) in order to help the institution in making better, data-driven decisions. The dashboards and interactive visualizations of the developed system will provide quick access to the extension activities’ real-time progress. The study is mainly focused on the business analytics of different operations and activities in extension services unit to systematize decision-making in order to upkeep real-time responses. During the development phase, it follows the dashboard development process [41].

The system employs a configurable Key Performance Indicators (Fig. 2) in developing the executive dashboards and interactive visualizations where extension services unit track and monitor their actual accomplishments on annually and quarterly basis. Since the system can be configured for all Philippine Association of State Universities and Colleges (PASUC) institutions, the researchers adopted the standard Major Final Output (MFO) for Extension Services of [42], particularly the key performance indicators, such as: (a) Number of trainees trained; (b) Number of trainees weighted by the length of training; (c) Number of extension programs organized and supported consistent with the SUC’s mandated and priority programs; (d) Percentage of beneficiaries who rate the training course/s and advisory services as satisfactory or higher in terms of quality and relevance; and (e) Percentage of persons who receive training or advisory services who rate timeliness of service delivery as good or better.

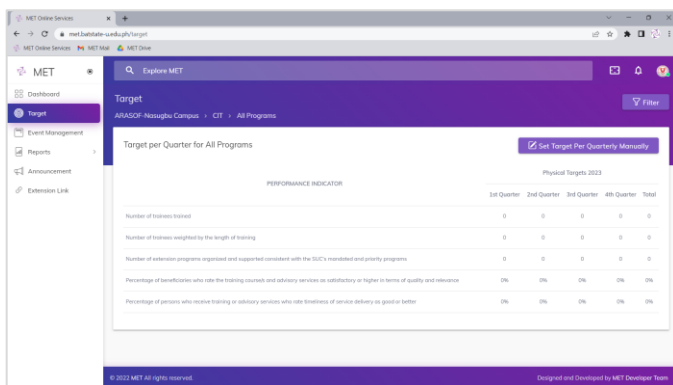


Fig. 2. Configurable key performance indicators

The executive dashboard and visualization features with filtering function [43] as shown in Fig. 3 to Fig. 5, provides the Extension Council, Institution and Academic Unit Heads along with the College Deans and Program Chairpersons with a high-level overview of the state of the extension services in their respective institutions/units/sectors. The system also generates beneficial reports on the extension services indicators which serve as a useful reference for the Office/Individual Performance Commitment Rating (OPCR/IPCR), Performance Based Bonus (PBB) and SUC leveling of the university.

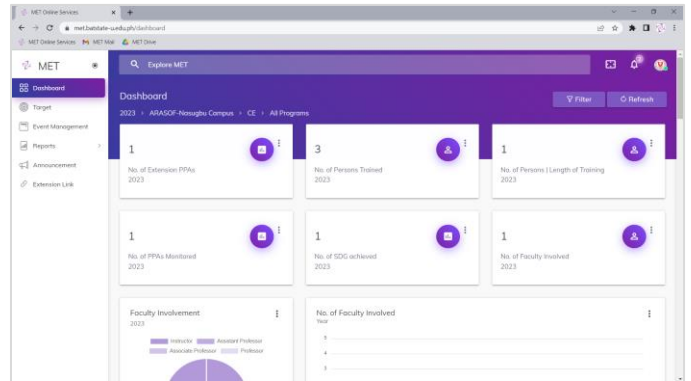


Fig. 3. Dashboards and interactive visualization (scorecards of extension key performance indicators)

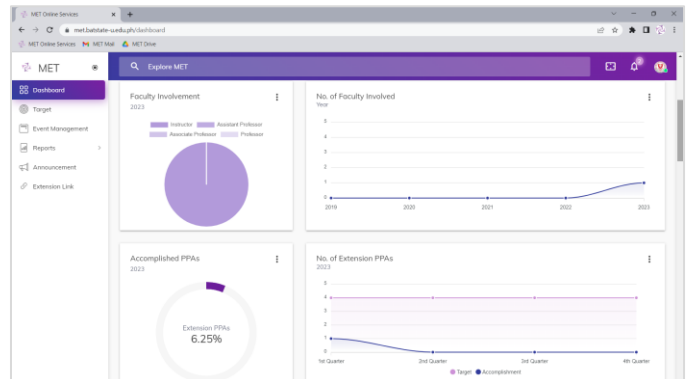


Fig. 4. Dashboards and interactive visualization (charts and graphs of faculty involvement and accomplished PPAs)

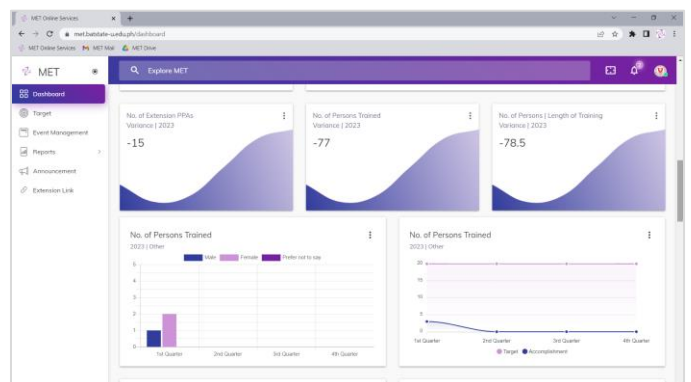


Fig. 5. Dashboards and interactive visualization (charts and graphs of variances and number of persons trained)

In addition to analytics part of the developed system, the following modules are incorporated to simplify the operations in the extension service unit: account management, target monitoring, event management, reports, announcement and extension links.

### A. Account Management

The Account Management module grants authorized users the rights to use the various services of the developed system by dividing them into groups and defining their privileges. The administration part of this module not only refers to the administration of the management system but also the administration tasks and information of the extension services unit task force.

Since MET Online Services is applicable also to other extension services unit of any State Universities and Colleges, the administration part of this module as shown in Fig. 6, is capable of configuring the following: information of the extension services unit, name of university with logo, campus/es, college/s, program/s, designation, employment status and academic rank of designated extension service task force, sustainable development goal, and type of community extension service.

Key Performance Indicators and evaluation questions and other settings are also configurable through the system administrator account. This is to fully utilize and accommodate different State Universities and Colleges (SUCs) with an end goal of improving their extension services, operations and transactions.

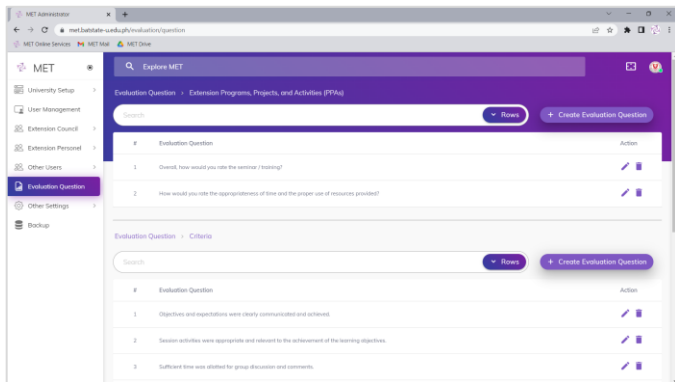


Fig. 6. Administrator account management.

### B. Target Monitoring

The Target Monitoring module involves managing and setting performance indicator targets. These transactions are processed by the institution and academic unit's extension heads along with designated college extension coordinators. The system provides a tool for the computation of key performance indicator in setting the target for the number of persons trained weighted by length of training with the corresponding legend for the weights of training (1) and percentage of beneficiaries who rate the training course/s and advisory services as satisfactory or higher in terms of quality and relevance, and timeliness (2).

$$a = b \times d \quad (1)$$

where,  $a$  is the number of persons trained weighted by length of training;  $b$  is the number of persons trained; and  $d$  is the weight of training.

Weights:

Length of Training	Weight
< 8 hours	0.50
8 hours (1 day)	1.00
2 days	1.25
3-4 days	1.50
5 days or more	2.00

$$y = (m/n) \times 100 \quad (2)$$

where,  $y$  is the percentage of beneficiaries who rate the training course/s and advisory services as satisfactory or higher in terms of quality and relevance, and timeliness;  $m$  is the number of trainees who rate the training course/s and advisory services as satisfactory or higher; and  $n$  is the total number of trainees surveyed.

### C. Event Management

The Event Management allows the institution and academic unit's extension heads as well as the college extension coordinators to manage all the extension services events / projects or activities. It includes electronic scheduling and task management of different extension activities. The status of each extension projects is automated and accessible on the web and mobile. There will be three (3) projects' status to wit: (a) For Implementation, (b) Ongoing, and (c) Implemented. These stages/statuses are dependent on the duration of the extension projects. It must be noted that once the project is implemented, it will be automatically added to the actual accomplishment. Different forms relevant to the extension services unit processes can be generated from this module, particularly the extension proposal, narrative report, monitoring proposal, and progress monitoring report.

### D. Reports

The Reports module is capable of generating monitoring and evaluation reports from the data entered from all the modules of the system including the android application used by the extension participants that evaluates the extension project. This module can be accessed and downloaded by the extension council, institution's extension head and academic unit's extension head, college dean / program chairperson and college extension coordinators (Fig. 7).

### E. Announcement and Extension Links

The announcement and extension links page will all the institution and academic unit heads to post and disseminate information to all extension services task forces relative to their regular functions and operations.

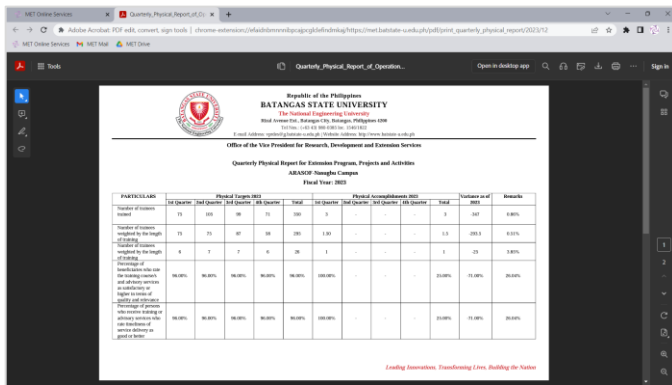


Fig. 7. Sample generated report using the prescribed template of the university

After the system development, a series of test cases and technical evaluations were carried out by a Certified Software Quality Assurance Specialist in order to ascertain, whether the developed system complies with the specified requirements. Software quality assurance (SQA) employs a methodical approach to identify trends and the necessary steps to enhance development processes. Unintended effects might result from finding and resolving coding problems; it is possible to solve one issue while also breaking other features and functions [44]. It can be gleaned from the test cases and technical evaluations as shown in Table I that the developed system is indeed feasible for implementation, proven to be fully-functional and passed the quality software standards set by the software quality assurance specialist.

TABLE I. TEST CASE AND TECHNICAL EVALUATION OF THE DEVELOPED SYSTEM

<p><b>Test Run Information:</b>                  Tester Name/s:                  A.Z. Leynes and J.R.M. Ruiz                  Date(s) of Test:                  September 9-13, 2022                  Location/server being used: N/A</p>	<p><b>Software Versions:</b>                  Application: MET Online Services                  Browser [used &amp; those COTS supports]: Google Chrome version 76.0.3809.132 (64-bit)                  Database: mySQL                  Operating System: Windows 10</p>
<p><b>Notes and Results:</b>                  Overall, we can say that this is a big help for all the extension coordinator. In this module you can monitor all the extension services key performance indicators with ease through dashboards and interactive visualizations. The user can also provide the status of each extension services what is the upcoming services, the on-going and implemented. The system can also provide reliable and consistent reports if each college coordinator hit the target accomplishment. On the other hand, the head of the extension services can also monitor all the college coordinator which is hassle free, the head of the extension coordinator will just login filter what college he/she can monitor what is the status of each college coordinator. And for reports the system provides printable reports that can use for documentation.</p>	
<b>Modules</b>	<b>Passed / Failed</b>
Account Management	Passed
Dashboard	Passed
Target Monitoring	Passed
Event Management	Passed
Reports	Passed
<b>Overall Evaluation</b>	<b>Passed</b>

As the system passed the software quality assurance, it is also clear as reflected from Table II that the developed system highly satisfied the users' expectations and requirements in terms of efficiency, functionality, reliability, usability, and security. The satisfaction of the end-users determines that the developed system is acceptable [45] and indeed feasible for implementation. It is also an indicator that the information system is effective when the users are satisfied with its functions and requirements [46, 47].

TABLE II. LEVEL OF SATISFACTION OF THE RESPONDENTS ON THE DEVELOPED SYSTEM

	Criteria	Weighted Mean	Verbal Interpretation
1	Efficiency	4.32	Highly Satisfied
2	Functionality	4.43	Highly Satisfied
3	Reliability	4.30	Highly Satisfied
4	Usability	4.57	Highly Satisfied
5	Security	4.32	Highly Satisfied
	<b>Overall Weighted Mean</b>	<b>4.39</b>	<b>Highly Satisfied</b>

The developed system automates the process of evaluation of the Extension Programs, Projects, and Activities (PPAs) as shown in Fig. 8. By using their mobile devices, the extension participants can evaluate online and raise their suggestions and comments based on the performance of the extension program conducted. It can be gleaned from Table III that the mobile application was highly accepted by the end-users in evaluating the extension programs, projects, and activities.

TABLE III. LEVEL OF ACCEPTANCE OF THE RESPONDENTS IN THE MOBILE APPLICATION OF EVALUATING THE EXTENSION ACTIVITIES

	Criteria	Weighted Mean	Verbal Interpretation
1	Ease of Use	4.30	Highly Acceptable
2	Usefulness	4.32	Highly Acceptable
	<b>Overall Weighted Mean</b>	<b>4.31</b>	<b>Highly Acceptable</b>

In case that the participants don't have mobile devices, the designated college extension coordinators and staff will be tasked to provide and assist the participants in evaluating the extension PPAs using the university's mobile devices or laptop provided by a unique evaluation code which is system generated. Another concern is the Internet connection, since the said developed information system required internet access for the evaluation, the researchers recommend that extension service office assisted by the Information and Communication Technology (ICT) office to provide laptop computers during the activity for the computerized evaluation and once there is already internet connection, the designated college extension coordinators will be the one responsible to sync the encoded evaluation answers to generate the evaluation results and update it online. Hence, there will be a fast and accurate evaluation results immediately after the online evaluation done by the extension participants.

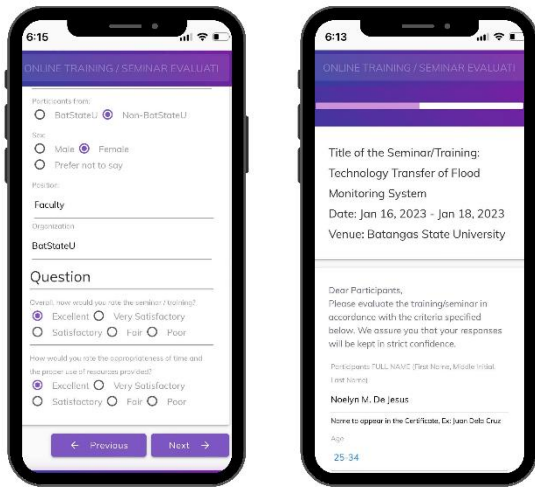


Fig. 8. Mobile evaluation of extensions activities.

## V. CONCLUSION

Most Extension Services Unit of SUCs used the manual process of managing, monitoring and evaluating their extension projects. This, most of the time, results in problems related to documents handling/archiving, timely monitoring and evaluation reports, poor management of budget allocation and utilization. The significance of this study is resolving those issues and challenges encountered in the extension services unit of the university by automating its operations and providing quick access to real-time progress of its activities through dashboards, interactive visualization, and reports.

The developed system is indeed feasible for implementation, proven to be fully-functional and passed the quality software standards of a Certified Software Quality Assurance Specialist. As the developed system satisfied the users' expectations and requirements, it would be an effective tool for the institution, extension services unit, and community, to make better strategic decisions and continuously deliver quality services to the community.

## VI. FUTURE WORK

It is suggested that future research should incorporate the evaluation for faculty involvement in extension which can be significant in the evaluation and promotion of faculty using the Qualitative Contribution Evaluation (QCE) instrument of the National Budget Circular (NBC) No. 461.

## ACKNOWLEDGMENTS

The authors would like to extend their profound gratitude and sincere appreciation to the Research, Development, and Extension Services Office, Batangas State University for supporting and funding this research. Also, to the Editors and anonymous reviewers for their valuable and constructive feedbacks to further improve the study.

## REFERENCES

[1] C.K. Pastor, "The Role of Management Information System: Review on the Importance of Data and Implementation in Organizational Process" 2020. [Online]. Available:

<https://ssrn.com/abstract=3558441> or <http://dx.doi.org/10.2139/ssrn.3558441>.

[2] A.O. Agbatogun, "Interactive digital technologies' use in southwest Nigerian universities" *Educational Technology Research and Development*, 61(2), pp. 333-357. [Online]. Available: <https://doi.org/10.1007/s11423-012-9282-1>.

[3] D.J. Amassoma, D.O. Ayanda and K. Tijani, "Internet usage among Nigerian polytechnic students and its impact on manpower development: A case study approach" *Journal of Information Technology Impact*, 10(3), 2010, pp. 161-170.

[4] QMS Solutions, "Management System Standards" QMS Solutions. [Online]. Available: <http://www.qmssolutions.com/quality-management-standards/>.

[5] J. Frankenfield, "Data Analytics: what it is, how it's used, and 4 basic techniques" *Investopedia | Financial Analysis*. 2022. [Online]. Available: <https://www.investopedia.com/terms/d/data-analytics.asp>.

[6] A.B. Shaqiri, "Management Information System and Decision-Making" *Academic Journal of Interdisciplinary Studies: Vol. 3 (No. 2)*. [Online]. Available: <https://doi.org/10.5901/ajis.2014.v3n2p19>.

[7] International Standard Organization, "Management system standards" ISO Standards. [Online]. Available: <https://www.iso.org/management-system-standards.html>.

[8] Batangas State University, "Extension Services Manual" Office of the Vice President for Research, Development and Extension, unpublished.

[9] I.I. Llenares and C.C. Deocariz, "Measuring the impact of an academe community extension program in the Philippines" *Malayan Journal of Learning and Instruction: Vol. 15 (No. 1)*, June 2018, pp. 35-55. [Online]. Available: <https://doi.org/10.32890/mjli2018.15.1.2>

[10] B.E. Swanson, R.P. Bentz, and A.J. Sofranko, "Improving agricultural extension. A reference manual" Food and Agriculture Organization of the United Nations Rome, 1997. [Online]. Available: <https://www.fao.org/3/w5830e/w5830e00.htm#Contents>.

[11] B. Davis, "What does extension mean in a lesson plan?" *Knowledge Bank*, 2021. [Online]. Available: <https://mv-organizing.com/what-does-extension-mean-in-a-lesson-plan/>

[12] M.G. Serrano, et.al., "An assessment of Community and Extension Service: Basis for Continual Improvement" *ASEAN Multidisciplinary Research Journal*, Vol 10(1), 2022.

[13] Commission on Higher Education, "Policy Standard To Enhance Quality Assurance (QA) in the Philippines Higher Education Through An Outcomes-Based And Typology-Based QA" CHED Morandum 46, series of 2012.

[14] N.I.D. Sermona, I. Talili, R. Enguito, and M.F. Salvador, "Implementation of Extension Services in select State Universities and Colleges in the Philippines", *Sci.Int.(Lahore)*, 32(6), 2020, pp. 609-614.

[15] J.S. Gannapao, "Rationalizing the Extension Programs of Abra State Institute of Sciences and Technology" *IAMURE International Journal of Ecology and Conservation*; 31(1). 2020. [Online]. Available: <http://ejournals.ph/form/cite.php?id=15303>.

[16] National Economic and Development Authority, "Philippine Development Plan 2017-2022" ISSN: 2243-7576.

[17] C.F. Gibbon, "Performance indicators" *BERA Dialogues (2)*, ISBN 978-85359-092-4.

[18] T. Weilkeins, C. Weiss, A. Grass, K.N. Duggen, "Frameworks" *OCEB 2 Certification Guide*. Elsevier. (2016) pp. 149-169. Available: doi:10.1016/b978-0-12-805352-2.00007-8.

[19] A.A. AlRababah, "A New Model of Information Systems Efficiency based on Key Performance Indicator (KPI)" *International Journal of Advanced Computer Science and Applications (IJACSA)*, 8(3), 2017. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2017.080313>.

[20] A. Abdelhadi, S. Zainudin, and N.S. Sani, "A Regression Model to Predict Key Performance Indicators in Higher Education Enrollments" *International Journal of Advanced Computer Science and Applications (IJACSA)*, 13(1), 2022. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2022.0130156>.

[21] J. Peral, A. Mate, and M. Marco, "Application of Data Mining techniques to identify relevant Key Performance Indicators. Computer



- Standards and Interfaces Volume 54, Part 2, November 2017, Pages 76-85, 2017.
- [22] A. Cadwell, "What is a Key Performance Indicator (KPI)? The Ultimate Guide," *Oracle NetSuite*, 2022. [Online]. Available: <https://www.netsuite.com/portal/resource/articles/erp/key-performance-indicators-kpis.shtml>.
- [23] KPI.org, "What is a Key Performance Indicator (KPI)?" [Online]. Available: <https://www.kpi.org/kpi-basics/>.
- [24] Wikipedia, "Performance indicator" [Online]. Available: [https://en.wikipedia.org/wiki/Performance\\_indicator](https://en.wikipedia.org/wiki/Performance_indicator).
- [25] A. Morris, "Descriptive Analytics Defined: Benefits & Examples", *Oracle NetSuite*, 2022. [Online]. Available: <https://www.netsuite.com/portal/resource/articles/erp/descriptive-analytics.shtml>.
- [26] J. Lismont, J. Vanthienen, B. Baesens, and W. Lemahieu. "Defining analytics maturity indicators: A survey approach" *International Journal of Information Management*, 2017. 37(3). pp.114–124.
- [27] K. Pauwels, T. Ambler, B. H. Clark, P. LaPointe, D. Reibstein, B. Skiera, B. Wierenga, and T. Wiesel. "Why, what, how, and what research is needed?" *Journal of Service Research*, 2009. 12(2): pp. 175–189.
- [28] D. M. Hanssens and K. H. Pauwels, "Demonstrating the value of marketing" *Journal of Marketing*, 2016. 80(6): pp. 173–190.
- [29] E. T. Bradlow, M. Gangwar, P. Kopalle, and S. Voleti., "The role of big data and predictive analytics in retailing" *Journal of Retailing*, 2017. 93(1): pp. 79–95.
- [30] M. Wedel and P. Kannan., "Marketing analytics for data-rich environments" *Journal of Marketing*, 2016. 80(6): pp. 97–121.
- [31] J. Bughin, "Ten big lessons learned from big data analytics" *Applied Marketing Analytics*, 2017. 2(4): pp. 286–295.
- [32] D. Delen and S. Ram. Research challenges and opportunities in business analytics. *Journal of Business Analytics*, 2018. 1(1): pp. 2–12.
- [33] O. Mintz, Y. Bart, P. Lenk, and D. Reibstein, "Drowning in metrics: How managers select and trade-off metrics for making marketing budgetary decisions" *SSRN Electronic Journal* 3502600, 2019.
- [34] C. Cote, "4 Types of data analytics to improve decision-making", *Harvard Business School*, 2021. [Online]. Available: <https://online.hbs.edu/blog/post/types-of-data-analysis#:~:text=Data%20visualization%20is%20a%20natural,a%20clear%20easily%20understandable%20way>.
- [35] K. Brush, and E. Burns, "What is data visualization and why is it important?", *SearchBusinessAnalytics*, 2020. [Online]. Available: <https://searchbusinessanalytics.techtarget.com/definition/datavisualization>.
- [36] ISBGlobal, "Whole-business reporting & Business Intelligence (BI)" *Business Intelligence*.
- [37] Wikipedia, "Dashboard (business)" *Wikipedia*. [Online]. Available: <https://en.wikipedia.org/wiki?curid=4166591>.
- [38] S. Few, "Information Dashboard Design: The Effective Visual Communication of Data" O'Reilly. 2006.
- [39] A. Watson, "Climb the Information Ladder" *Medium*. 4 Sept. 2020.
- [40] O.M. Yigitbasioglu and O. Velcu-Laitinen, "The Use of Dashboards in Performance Management: Evidence from Sales Managers" *The International Journal of Digital Accounting Research*. 2012. 12: 36-58. [Online]. Available: [http://dx.doi.org/10.4192/1577-8517-V12\\_2](http://dx.doi.org/10.4192/1577-8517-V12_2).
- [41] "Dashboard Development Process", *Unilytics*, 2016. [Online]. Available: <https://unilytics.com/data-visualization/dashboard-development-process/>
- [42] Commission on Higher Education, "Supplemental Implementing Guidelines on Cascading Performance Targets of State Universities and Colleges (SUCs) in line with Executive Order (EO) No. 80 Series of 2012" CHED Morandum 29, series of 2013.
- [43] N.I. Ismail, N.A.S. Abdullah and N. Omar, "Exploring Alumni Data using Data Visualization Techniques" *International Journal of Advanced Computer Science and Applications (IJACSA)*, 13(9), 2022. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2022.0130922>.
- [44] A.S. Gillis, "Quality Assurance (QA), ITech Target Network. [Online]. Available: <https://www.techtarget.com/searchsoftwarequality/definition/quality-assurance>.
- [45] R.K. Sinha and S. Kurian, "Assessment of end user satisfaction of hospital information system" *Management in Health*. 2014; 18(3).
- [46] J.Y.L. Thong, C-S Yap, and K.S. Raman, "User Satisfaction as a Measure of Information System Effectiveness" In: Stowell FA, West D, Howell JG, editors. *Systems Science: Addressing Global Issues*. Boston, MA: Springer US; 1993. p. 487-92.
- [47] V.P. Aggelidis and P.D. Chatzoglou. "Hospital information systems: Measuring end user computing satisfaction (EUCS)" *Journal of biomedical informatics*. 2012 Jun 1;45(3):566-79.

# Arabic Stock-News Sentiments and Economic Aspects using BERT Model

Eman Alasmari<sup>1</sup>, Mohamed Hamdy<sup>2</sup>, Khaled H. Alyoubi<sup>3</sup>, Fahd Saleh Alotaibi<sup>4</sup>,

The Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia<sup>1,3,4</sup>  
The Faculty of Computer and Information Sciences, Ain Shams University, 11566 Cairo, Egypt<sup>2</sup>

**Abstract**—Stock-market news sentiment analysis (SA) aims to identify the attitudes of the news of the stock on the official platforms toward companies' stocks. It supports making the right decision in investing or analysts' evaluation. However, the research on Arabic SA is limited compared to that on English SA due to the complexity and limited corpora of the Arabic language. This paper develops a model of sentiments to predict the polarity of Arabic stock news in microblogs based on Machine Learning and Deep Learning approaches. Also, it aims to extract the reasons which lead to polarity categorization as the main economic causes or aspects based on semantic unity. Therefore, this paper presents an Arabic SA approach based on the logistic regression model and the Bidirectional Encoder Representations from Transformers (BERT) model. The proposed model is used to classify articles as positive, negative, or neutral. It was trained based on data collected from an official Saudi stock-market article platform that was later preprocessed and labeled. Moreover, the economic reasons for the articles based on semantic unit, divided into seven economic aspects to highlight the polarity, were investigated. The supervised BERT model obtained 88% article classification accuracy based on SA, and the unsupervised mean Word2Vec encoder obtained 80% economic-aspect clustering accuracy.

**Keywords**—Machine learning; deep learning; classification; prediction; statements

## I. INTRODUCTION

Stock price decline is a concern for stock market investors and an exciting issue for stock analysts. Stock market news sentiment analysis (SA) allows investors to make the correct investment decision through thorough and thoughtful deliberation. Whereas knowing the best time to buy or sell stocks is the main aim of stock market prediction as there are various factors that may affect stock prices, these factors depend on stock market news indicators, which economists take advantage of to analyze the factors that may affect the market, such as public news sentiment [1]. Undoubtedly, everyone is familiar with sharing news with others in social communities either individually or as a company [2]. Currently, online microblogs' news makes vast amounts of news moods available. At the same time, microblogs may contain large amounts of data about companies that share their information, such as that about their stocks and the stock market, through their microblogs. The articles on these microblogs are used to classify their stocks' orientations and the economic causes of these.

Analysts and investors still have difficulty predicting the future price of a company's stocks. The sentiment of news

regarding a company's stocks is a vital interest of financial analysts, investors, and other competitors [3, 4]. SA aims at sentiment recognition and public-opinion checking, which are considered text-mining research fields [5]. Although there are many types of research for analyzing individuals' behaviors in social microblogs, few studies have analyzed the classification of stock news in financial-market microblogs [6]. These microblogs allow investors and analysts to share financial news and opinions with other investors. The SA of microblog posts helps investors perceive the risks related to companies' stocks, which assists them in their investment decision-making investor sentiment shown in microblogs has a significant impact on stock prices relative to growth stocks [7]. Therefore, the prediction of stock market behavior depends on the polarity prediction classification of microblogs' news sentiments. Such news can be categorized as positive, negative, or neutral for the polarity prediction of SA, which affects the prediction of stock news classification [3, 4, 8, 9]. Using companies' information from microblogs can therefore improve the accuracy of polarity prediction. Thus, there is a need for SA of companies' stock market articles for stock price prediction.

The previous studies on SA in deep learning (DL) and machine learning (ML) heavily focused on analyzing several features and ignored the sentences' structures and the relations between words in any targeted language, such as the Arabic language [10, 11]. Also, the recent studies have focused on applying DL in SA for multiple tasks based on the Arabic language [11]. However, to the best of the authors' knowledge, few of the previous DL studies were on the Arabic SA in the stock market fields. Moreover, there has been no study that bridges the gaps in Arabic SA by using DL with modern standard Arabic (MSA) features in the Arabic sentiments mirrored in stock news even though DL SA approaches such as the Bidirectional Encoder Representations from Transformers (BERT) model currently offer the best performance for Arabic stock news [12, 13]. Besides, no study on the extraction of the main economic aspects of the reasons for the sentiment polarity depending on Arabic news features has yet been conducted. Thus, there is a need to develop ML and DL models that can classify stock news sentiments and determine their polarity on the basis of labeled data, and that can extract the economic causes of such sentiments. The current study focused on the prediction and classification of Arabic stock market news sentiments and their main economic causes. It sought to develop a supervised model based on logistic regression and the BERT classifier for labeled Arabic data, and an unsupervised model based on k-means clustering.

## II. RELATED WORK

### A. Arabic Sentiment Analysis Techniques for Microblog News

The Arabic SA models and techniques consist of steps such as preprocessing and analysis of linguistic, part-of-speech (POS), semantic, and lexicon-derived features [14].

Preprocessing is often the first step in a text-processing system. It consists of steps for facilitating classification. For instance, in Arabic text preprocessing, the text is tokenized; the letters at the beginning of names (e.g., “ال”) are removed; the letters are normalized, such as by converting “أ” (“Hamza”) to “ا” (“Alef”); and then the stop words, such as “كان, في, إن” (“was, that, in”), are removed [15]. Thus, preprocessing cleans the textual data by removing the undesirable elements therefrom to increase the accuracy of the future SA results. Ignoring preprocessing such as spelling corrections will make systems disregard the main words [14, 16, 17], but overdoing preprocessing will make systems lose important data.

Preprocessing of microblogs increases the sentiment prediction accuracy in any field. It considers the variations of letters, such as the basic Arabic letter “Alef” or “أ,” which includes the derivative letter “Alef Maksoura” or “آ” often confused with the other basic Arabic letter “Ya” or “ي.” Also, there is usually confusion in writing the basic Arabic letters “Ta marbota” or “ة” and “Ha” or “ه.” Moreover, the letter “Hamza” is an interchangeable letter based on its word and position, which include “أ”, “ؤ”, and “ئ.” However, additional preprocessing approaches need to be used for some microblog data. These additional techniques involve several tasks, such as the elimination of hashtags, mentions, URLs, and special characters and reposting after the elimination of these elements [14]. Thus, it is essential to handle the Arabic dialects or to provide an MSA alternative.

Linguistic-feature analysis, the second step in the Arabic SA models and techniques, is divided into two techniques: analysis of n-grams and analysis of the POS features. N-gram is a chain of n-elements from some textual data, such as words, letters, and syllables. The frequently used n-gram elements are words, which are categorized into three types: unigrams (one word), bigrams (two sequential words), and trigrams (three sequential words) [14, 16, 17]. In Arabic SA, unigrams lead to a high prediction performance, along with the syntactic features. Syntactic features such as word roots, word n-grams, and punctuation impact SA. The variations in the Arabic syntactic features lead to the use of the word roots because although Arabic words have many forms, they all originate from the same root. For instance, all the Arabic words “سلمت”, “يسلم”, and “سلام” come from the root word “سلم.” Thus, there are two configuration settings for extracting the roots of Arabic words: lexeme (LEX) and lemma (LEM). LEX is a configuration setting of all word forms containing the same meaning through tokenization and morphotactics. LEM, on the other hand, is the exact form selected to represent LEX [14]; thus, the Arabic noun is the masculine singular default form but the verb is the third-person masculine singular perfective.

POS tagging consists of dividing a word into grammatical categories: nouns, pronouns, verbs, adverbs, adjectives, conjunctions, interjections, and prepositions. The POS tags in Arabic-text analysis include information about the morphology of the word. As for the semantic features, they contain contextual features that point to the semantic orientation of the surrounding textual data. Grammatical features have annotation approaches that add a polarity score to phrases or words by measuring the overall correlation of a multilabel of elements. This semantic orientation is based on various concepts of elements with a given sentiment polarity. Hence, if this element did not appear in the prior dataset with the association of the larger group of elements, polarity can be detected. Some researchers have found that semantic features outperform the unigram and POS-tagging features. Finally, the lexicon-derived features of the Arabic language are lexicons automatically created from some microblogs. However, Arabic lexicons are different from English lexicons as the Arabic language has various dialects and various word forms originating from a single root word [14].

### B. Targeted News Data

1) *Importance of the news data in microblogs:* Microblogs have become essential channels for news dissemination. An increasing number of users are expressing their feelings and sharing other information about the social news in microblogs [18]. Therefore, microblogs allow investors to share financial news and opinions with other investors. SA of microblog posts helps investors perceive the risks involved in buying a company’s stocks and make a decision regarding whether to invest in a company or not to. Hence, stock market microblog news SA has a significant impact on the stock price relative to growth stocks and is associated with stock price movement [7, 10, 16]. The decision-making regarding buying/selling orders is based on the market mood determined by technical indicators, which are measurements based on stock prices’ time series [19]. Also, unexpected events related to companies affect their performance in a positive or negative direction, such as their stock prices [10]. Consequently, for investment purposes, stock price forecasting has attracted increasing attention in recent years [20], and microblog news data may be among the essential inputs for such forecasting [21]. Thus, stock market prediction values have a significant impact on the financial sector.

Stock price prediction is important due to the high volatility of stock market prices [22]. Therefore, to minimize the risk of stock market investment, an accurate stock market price prediction model is needed. The investor sentiment index has been used for stock market prediction in recent years. It is based on stock market news, which some vendors provide as a service. To avoid buying overrated or high-risk stocks, investors decide whether to buy or sell stocks depending on the main stock news. Stock prices and their movements are thus forecasted by analyzing the related news, which outperforms an investor’s forecasting of the upcoming short-term trends [10, 23, 24]. Thus, automated decision-making supports the prediction of the upcoming stock price trends.

2) *Challenge of obtaining news data:* The primary challenge faced by researchers with regard to data collection from microblogs is the recent changing of some microblogs' terms of service, disallowing public hosting of old textual data or blogs and public extraction of these [16]. Besides, microblogs have increased the number of orders on API to retrieve specific data, such as on Twitter, as they can no longer directly obtain many data resources. Thus, microblog crawling tools are used to obtain many data resources. There are specific crawling processes: (1) selecting the number of microblogs with particular features, such as language; (2) crawling for a specific type of microblog data and filtering the textual microblog after obtaining each piece of data; (3) archiving the data obtained from crawling on the basis of the ID of each microblog in a specific period; and (4) processing the microblog texts by removing the miscellaneous elements, extracting the themes and feelings, segmenting the words, analyzing the syntax, and processing the text [18]. Accordingly, the Arabic microblogs of some news websites that use a crawler to extract data contain many labels within a specific period [15].

The user-written textual dataset is usually a massive volume of noisy and unstructured data. These data make SA a difficult and challenging task; thus, there is a need to process the unstructured data computationally to extract and determine the sentiments embedded in them [24]. Each language used to write microblog contents has a unique form or word structure [18]. The sentiment of the microblog is affected by the word sequence features, textual-language features, and grammatical-relation features [18]. These challenges can be dealt with through feature filtering (by analyzing some feature sets, such as n-grams) and by replacing certain words and processing the text, which become more significant as the data increase [16].

The processing of textual data involves removing any word that is irrelevant to the text's sentiment. Therefore, the preprocessing of such textual data involves scoring the text's sentiment after excluding the noise therein caused by the aforementioned words. Text processing is done using Python's Natural Language Toolkit on the basis of two methods: tokenization and removing stop words and symbols. Tokenization involves dividing texts by spaces to make a list of individual words per word package. Each word is then used as a feature to train or learn the classifier. Stop words are removed from the list of words because they have neutral meanings and are inappropriate for SA in the targeted language, according to its dictionary. Stop words may include prepositions, symbols such as "@" and URLs [16], and other words that have no sentiment value.

Unigrams are one-word n-grams for each unique tokenized word made for the classifier. For instance, a microblog can be classified according to whether it contains the word "bad" or does not. As this unigram is associated with a negative microblog, the classifier will classify microblogs containing the word "bad" as negative. Bigrams (two-word n-grams) and trigrams (three-word n-grams) can also be classifiers; they classify microblogs on the basis of whether they contain two or three words, respectively, or do not [16, 17]. For example,

if a microblog contains the bigram "not bad," which is associated with a positive microblog, it will be classified as a positive microblog.

Finally, word replacement involves replacing a company's stock symbols and positive and negative words.  $N$  most similar words according to a cosine similarity can be used to replace negative and positive words. Stock symbols can be replaced with a common word and can be removed from the textual blog [17], (e.g., \$AAPL can be replaced with "company").

### 3) *Sentiment classification approaches*

a) *Neural network models:* The artificial neural network (ANN) models have achieved high performance in Arabic classification and prediction on the basis of the related work chapter. The ANN has neurons in each layer. Each neuron calculates each input and the sum of the weights, adds the bias, and performs the activation function (e.g., the sigmoid function). The ANN model develops algorithms to solve complex problems such as prediction problems. Neural network models can be generalized by learning from the inputs' relationships to predict the unseen relationships on new data. This enables the model to predict and make generalizations regarding the unseen data. Neural-network prediction approaches perform better with high-volatility data by learning the unobserved relationships in the data without setting any specific relationships in the data. Thus, neural-network models are useful in financial time series prediction [25, 26], such as in the prediction of stock prices, which contain high-volatility data.

b) *Deep-learning model:* The DL techniques are used in SA because of their high performance in prediction [27]. They can yield a high result in financial applications. Their impact depends on training complex nonlinear models on the basis of massive datasets [19]. There have been studies on SA approaches based on DL for accuracy improvement, but some DL techniques ignore the words' meanings and order. Therefore, recurrent neural networks (RNNs) such as long short-term memory (LSTM), gated recurrent unit, and BERT are used to train models with solid architectures. The RNNs' architectures extract features in sequential and non-sequential data [28, 29]. The DL approaches obtain the highest results in the binary-data strategies. Thus, the previous studies' continuous-data outcomes showed that RNN and LSTM have the best classification performance [30]. Also, the sentiment classification by Bidirectional Encoder Representations Transformer-Bidirectional models such as BERT and BERT-BiLSTM [31], are superior in accuracy to other classifiers.

The BERT model takes advantage of bidirectionality and adds a masked language model (LM) to hide the word predicted and for next-sentence prediction (NSP) [32, 33]. Using such model addresses the limitations of the previous LMs, which work from left to right and do not capture the bidirectional contexts. Also, such model generalizes LM for easy fine tuning for any downstream task. The BERT model mainly performs two steps in its framework: pre-training and fine tuning. Pre-training is done through a couple of unsupervised tasks: masked LM and NSP. The BERT model

adds two tokens: Class (CSL) and two Separation Sentences (SEPs) for the input sequences as a separate structure feature and target [33], to be used in one or two sequences.

### III. MATERIALS AND METHODS

The proposed model determines the classification polarity of data and the economic-reason polarity through data gathering, data preparation, and data splitting. Then the supervised and unsupervised models are applied.

#### A. Data Gathering

The data that were used in the present study were collected from the Saudi stock market platform Tadawul, containing the Corporate Articles and Historical Data Stocks datasets. The total dataset covered articles published within nine years (2011–2019). It had 34,386 rows of news and 16 columns of variables: sector, investor name, investor ID, date, time, article title, full article, opening value, highest price, lowest price, closing value, change %, change value, quantity handled value, total price, and total of deals.

#### B. Data Preparation and Annotation

In the proposed approach, there are multiple phases of data preparation: data evaluation, data cleaning, and data validation. Some issues (e.g., noisy, incorrect data and differentiated data formats) are fixed in the data evaluation stage. The data-cleaning stage handles these issues. Noisy and incorrect data are filtered and erased manually and by Pandas DataFrame. Arabic tokenization is performed to fix the differentiated data formats and to standardize all the data sections. In the end, a manual test is performed to validate the whole dataset by testing 30,089 observations.

The polarity annotation adopts the multi-class classification system, which includes the neutral class to increase the model's prediction accuracy [34, 35]. Stock news polarity is annotated manually as negative, positive, or neutral by native Arabic speakers associated with the stock market field.

#### C. Data Splitting

Some data-splitting methods, such as k-fold, grid search, and cross-validation, split the dataset into a couple of sets. These methods are correct but take a long time to train and evaluate each hyperparameter value, which is ineffective in DL models. Therefore, splitting the dataset into three subsets (train-validation-test splitting) increases the training and evaluation speed. Train-validation-test splitting chooses the best hyperparameters by evaluating the validation set outcomes after the training level. Then it re-evaluates the test set outcomes after passing the validation level [36]. Thus, train-validation-test splitting allows selecting the best model, features, and hyperparameters on the validation and test levels. It decreases the mistake cost and performs DL fast for long-time training and evaluation, to the best of the authors' knowledge.

#### D.

#### The Applied Models

Two models were applied in the current study: the supervised DL model of SA classification and the unsupervised DL model of economic-aspect clustering. Stock news sentiment classification was performed using the BERT model; the articles were classified as positive, negative, or neutral. The unsupervised cluster was improved by performing semantic k-means clustering to categorize the articles into main economic aspects. Thus, these models classify the stock news polarity on the basis of labeled data and extract the economic reason for each stock news article's sentiment.

4) *Supervised model*: Two models were applied to the polarity classification, as shown below.

- Baseline model (BL): Logistic regression
- DL model: BERT

a) *Logistic regression model*: Starting with a simple BL such as the logistic regression model before the complex models helps test the data quality, estimate the primary result, and determine the problem dimensions. The logistic model works effectively with multi-class classification and can be generalized [37]. Also, it performs better with a massive amount of data, such as the dataset used in the present study [38]. The steps below are applied to obtain the best prediction result.

- The data are lemmatized to remove the noisy features.
- Trigrams add the important features.
- The term-frequency-inverse document frequency (TFIDF) vectorizer is used to give the essential words more weight and vice versa.
- The maximum number of features is limited according to the best result.

Weight balancing can be learned using the stochastic gradient descent (SGD) algorithm; thus, the alpha value equals 0.0001 because the TFIDF is the one that controls the importance of features to give proper weights to the important words. The logistic regression model's performance is tested on the basis of the max\_features value, by changing the size of the vocabulary. Therefore, many values are tried, and the results decrease at 20,000. Thus, 25,000 is the best choice, with the best result and the fewest maximum features. The maximum number of features is limited using the argument "max\_feature" by the fewer features and best result simultaneously. The multi-classes in the training stage are balanced using the argument "class weight" to prevent the imbalanced-learning effects on the classification performance. Table I shows the parameters' values which are setted to perform logistic regression on multi-class classification, by tuning the best result.

TABLE I. THE LOGISTIC REGRESSION MODEL MULTI-CLASS CLASSIFICATION PARAMETERS

Model	Parameter	Value
Logistic Regression	“CountVectorizer ngram_range”	1, 2
	“CountVectorizer max_features”	250,000
	“SGDClassifier __loss”	“log”
	“SGDClassifier __n_jobs”	-1
	“SGDClassifier __class_weight”	“balanced”
	“SGDClassifier __alpha”	0.00001
	“SGDClassifier __random_state”	1

b) *The bidirectional encoder representations from transformers model:* BERT comes up with a general understanding to be used in a downstream task for text classification. It trains a strong model language and then adds an output layer in each downstream task of such language. This output layer is suitable for the downstream task, allowing the fine tuning of this layer and of the other layers [33]. Using the BERT model along with WordPiece embedding will solve this problem. WordPiece takes care of the embedding for the whole token, and the embedding also divides pieces of it. Thus, the model does not discard the newly faced word in the test data but takes the embeddings of the token pieces in the training data [33, 39]. An example is if the model faced the term “going” in the training data. In the previous models, the word “going” is a full token that has only one embedding. Thus, if the previous models faced the new tokens “go” and “ing” in the test data, they would discard such tokens.

However, WordPiece gives embeddings for “go” as the first piece and “ing” as the second piece. Thus, with WordPiece, when new tokens like “go” and “ing” are faced in the test data, both will be predictable on the basis of the set embeddings’ values, and there is no need to discard them.

Table II shows the fine tuning of the optimal hyperparameter values is task specific. Still, the ranges of possible values for the following are studded to work well across all tasks on the basis of the BERT study [33]: batch size, learning rate, and number of epochs, where the best epoch equals 4 in the tests and experiments. The maximum length of documents is chosen on the basis of the quantile value, which equals 0.9; that is, 90% of the documents are less than or equal to 425 words, as the articles’ maximum length (see Fig. 1). The Max\_features value equals 30,000 based on the previous experiments, and BERT model authors [33]. The wall time of training equals 1 minute and 3 seconds.

TABLE II. FINE-TUNING PROCEDURE OF THE BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS MODEL

Model	Parameter	Value
BERT	Batch size	16, 32
	Learning rate (Adam)	5e-5, 3e-5, 2e-5
	Number of epochs	2, 3, 4
	Classifier (Dense)	2,307
	Quantile	0.9
	Best_epoch	4
	Max-features	30,000

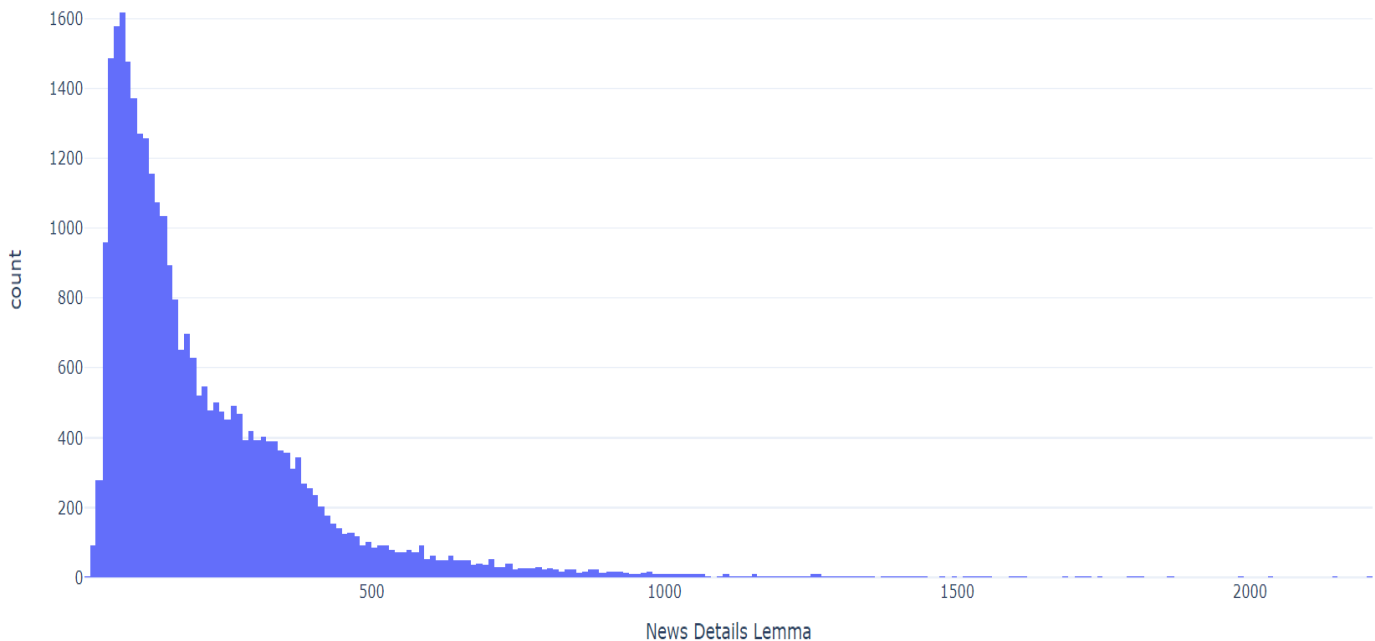


Fig. 1. Maximum length of documents based on the quantile value

BERT prediction of multi-class classification was applied to compare the results of the validation set with the actual target data. Table III shows the evaluation outcomes of the data whose classifications were predicted.

TABLE III. MULTI-CLASS CLASSIFICATION PREDICTION RESULTS OF THE BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS MODEL

	Negative	Neutral	Positive
Negative	0.8842	0.0450	0.0707
Neutral	0.0218	0.8667	0.1115
Positive	0.0277	0.0748	0.8975

5) *Unsupervised model*: As shown in the previous sections, the BERT model was trained to predict the correct polarity of stock market news articles. However, although this is useful by itself, it misses some parts of the needed picture. The preceding section showed how the polarity of an article is determined but does not elaborate on why the article carries such polarity. For example, the BERT classifier classifies an article as positive, but the reason for its classification of the article as positive is unknown. The reasons for polarity can include the company profits and the fact that the company is to start a new project or production line, increase its capital, or embark on a merger or acquisition. Such assessment is invaluable to the investors, the company owners, and the company's board of directors. The determination of the economic reason for an article's polarity in the literature is called aspect-based SA, which is performed through the procedure described below.

- Annotating each article according to the aspect involved and the underlying sentiment
- Training a supervised ML or DL model on the annotated data
- Using the trained model to predict the aspect and sentiment of new articles

The aforementioned procedure, however, is extremely time consuming and costly. Therefore, the approach proposed herein is a novel and universal approach that significantly reduces both time and cost by reformulating the problem of aspect extraction from a supervised text classification to an unsupervised determination of text similarity, hence eliminating the hand-annotating step from the typical approach described earlier. The proposed approach is performed through several steps. The first step is discovering the overall prevailing categories of aspects in the entire body of articles. Semantic k-means clustering is applied on the most important features per the trigram-TFIDF logistic regression model [40]. Also, instead of performing k-means clustering on the full articles, k-means clustering is limited to the reasons for the negative or positive polarity, as understood by the model. The second step is representing each category of aspects through a set of n-grams carrying the distinct semantics of the category. The third step is encoding the n-grams for each category into an embedding using a sentence-embedding procedure (transformer-based multilingual universal encoder and the mean Word2Vec encoder). Next,

the articles are encoded using the same encoder chosen in the third step. Instead of encoding the whole article at once, the article is encoded after paying attention to the relationships between the words and after aspect embedding for improving the goals. Then, for each article, the arc-cosine similarity between the embedded articles is computed, and each of the category embeddings in the third step is computed. Also, a score from 0 to 1 is given for how much each category is related to the article. Finally, the two aspects most associated with the highest scores are picked. Thus, the number 2 is a hyperparameter picked on the basis of reasoning, as will be explained later.

K-means clustering uses the mean Word2Vec vectorizer and needs to choose the number of clusters required by the algorithm for separating the features into. In the present study, after many possibilities were experimented with, ten groups were found to provide good semantic separation between the aspect categories. Hence, some overall categories of reasons (or aspects) were detected for positive and negative sentiments on the basis of k-means clustering. The following are the ten clusters that resulted from k-means clustering using the mean Word2Vec vectorizer based on the semantic unit.

- Clusters 0, 1, and 3 contain articles about profit/revenue increase/decrease, including stock appreciation/depreciation.
- Cluster 4 contains articles about starting/stopping a project or a production line, or anything related to production in general.
- Cluster 6 contains articles about the increase/decrease of the company's capital.
- Cluster 7 contains articles on dividend distribution or non-distribution.
- Cluster 8 contains articles on insurance operation surplus/shortage.
- Cluster 2 contains articles on the approval of various insurance policies.
- Cluster 9 contains articles on the board of directors' signing of agreements/memoranda of understanding, approval, or disapproval on critical issues or on anything related to managerial matters.
- Cluster 5 has miscellaneous articles that have been not considered.

Most of the clusters exhibited sentimental unity rather than semantic unity, which was the present study's aim for the economic aspects. The Universal Sentence Encoder was also applied to cluster the articles into ten groups. However, unlike the mean Word2Vec encoder, it cannot separate the semantics from the sentiments. The mean Word2Vec encoder is thus the best choice for exhibiting only semantic unity, as required. The central economic aspects manually extracted on the basis of the previous results were profit/revenue, projects/production lines, capital, dividends, insurance operations, insurance approval, and managerial/contracts.

TABLE IV. WORD2VEC VECTORIZER AND ATTENTION ALGORITHM  
EXAMPLES OF TESTING

Article 1 (Arabic)	Article 1 (English)
<p>إشارة للإعلان السابق للشركة المتقدمة للبتر وكيمياءات المتقدمة المنشور على موقع تداول بتاريخ 25 أكتوبر 2015م بخصوص اكتمال الاعمال الميكانيكية لمشروعها المشترك الخاص بمصنع انتاج البروبلين بجمهورية كوريا الجنوبية بطاقة تصميمية تبلغ 600 000 طن متري في السنة تود المتقدمة ان تعلن انه بتاريخ اليوم الثلاثاء 15 مارس 2016م بدأ مصنع انتاج البرولين المملوك لشركة اس كي ادفانسد المحدودة اس كي ادفانسد عملية التشغيل التجريبي للانتاج الذي سيخضع لاختبار الاداء طبقا للعقدي ترخيص التقنية مقابل اعمال الهندسة التوريد والتشييد الذي قد يستغرق من ... شهر الي ثلاثة اشهر</p>	<p>This is a reference to the previous announcement of Advanced Petrochemical Company published on the Tadawul website on October 25, 2015 regarding the completion of the mechanical works for the company's joint project for the propylene production plant in the Republic of South Korea, with a design capacity of 600,000 tons per year. On Tuesday, March 15, 2016, the applicant announced that the propylene production plant is owned by SK Advanced Ltd.; the process of trial operation of the production, which will undergo a performance test according to the technology licensing contracts; the engineering works contractor; and that the construction may take from 1 to 3 months...</p>
<b>Normal Algorithm Result</b>	<b>Attention Algorithm Result</b>
<p>1. Insurance approval: 0.8260993022742971 2. Profit: 0.79493785 3. Production: 0.73466593 4. Dividends: 0.70566356 5. Managerial/contracts: 0.6876711 6. Insurance ops: 0.6737051 Capital: 0.6472212</p>	<p>1. Insurance approval: 0.8508865377472663 2. Profit: 0.8629385 3. Dividends: 0.77983284 4. Production: 0.7792212 5. Managerial/contracts: 0.723952 6. Insurance ops: 0.7149271 Capital: 0.67267907</p>
Article 2 (Arabic)	Article 2 (English)
<p>تعلن الشركة السعودية لخدمات السيارات المعدات ساسكو انها بتاريخ 13 ديسمبر 2015م قد اكملت توقيع اتفاقية تسهيلات مرابحة متوافقة مع الشريعة الاسلامية مع بنك الخليج الدولي شركة مساهمة بحرينية بقيمة 150 مليون ريال ذلك بضمان سند لامر تتضمن هذه الاتفاقية قرض متوسط الاجل بقيمة 50 مليون ريال بمدة تمويل خمس سنوات منها سنتين فترة سماح علي ان يتم سداهه من خلال اقساط ربع سنوية متساوية القيمة بالإضافة الي اصدار خطابات ضمان بقيمة 100 مليون ريال يكمن الهدف من راء هذه الاتفاقية التوسع في مشاريع الشركة دعم انشطتها الرئيسية شراء مواقع جديدة لبناء محطات قود فضلا عن تمويل راس المال العامل ...</p>	<p>The Saudi Automotive Equipment Services Company (SASCO) announces that on December 13, 2015, it completed the signing of a Murabaha facility, which is compatible with Islamic Sharia with Gulf International Bank, a Bahraini joint-stock company worth 150 million riyals, by guaranteeing a bond. This agreement includes a medium-term loan of 50 million riyals with a financing period of 5 years, including a 2-year grace period, provided that it is repaid through quarterly installments of equal value. In addition to issuing letters of guarantee amounting to 100 million riyals. The objective behind this agreement is to expand the company's projects, support its main activities, purchase new sites on which to build fuel stations, and finance working capital, ....</p>
<b>Normal Algorithm Result</b>	<b>Attention Algorithm Result</b>
<p>7. Profit: 0.7810267 8. Insurance approval: 0.7601846408453664 9. Production: 0.74592507 10. Dividends: 0.70682096 11. Insurance ops: 0.702791 12. Managerial/contracts: 0.68651867 13. Capital: 0.6522217</p>	<p>1. Profit: 0.82416815 2. Insurance ops: 0.85664684 3. Insurance approval: 0.7774521435437018 4. Production: 0.7726567 5. Dividends: 0.75157106 6. Managerial/contracts: 0.73063576 Capital: 0.71698725</p>

Testing the proposed approach using individual articles showed a large percentage for each aspect based on the article's relevance to each aspect. There is testing for the applied algorithms: k-means clustering using the mean Word2Vec vectorizer and attention algorithm. The attention mechanism enhances the encoding of the article instead of treating the constituent words in the article equally. Attention is a method of selectively encoding an article by assigning higher weights to the words that are semantically similar to each corresponding aspect. However, although the two algorithms' results are similar, the normal algorithm has more sensible results in many examples. Table IV contains two tested examples for both algorithms. The first article is about insurance and production lines; the normal algorithm accurately predicts its aspects. Also, the attention algorithm is close to the result, but not to the best one. The second article is about profit, insurance approval, and production. Both algorithms accurately predict the aspects, but with some differences in the arrangement. Thus, it is essential to detect the highest aspect or two aspects to limit the probabilities and to validate the method performance. Thus, the proposed approach predicts the highest uni-aspect and bi-aspects besides the seven aspects' percentages of similarity to the article.

The results look promising but not perfect. The algorithm tends to favor the dividends aspect more than the other aspects, although it does this only when the article talks about profit or loss or financial statements as dividend distribution or non-distribution is closely tied to how a company is faring in the financial aspect. Even when the algorithm deviates from the correct aspect, it favors the closely correlated aspects.

#### IV. RESULTS AND DISCUSSION

##### A. Polarity Prediction

After the performance of many experiments, the trigram algorithm generated high results but a massive number of features. Therefore, a trigram is used with TFIDF to maintain the performance and limit the number of elements simultaneously. Also, the highest number of features is specified using the argument "max\_feature=250000," resulting in fewer features. It chooses features on the basis of their importance, by limiting the features with the best results. Table V includes the logistic regression and BERT model performances which are assessed through model testing and evaluation of the classification by the best model. Model testing requires feature and target data. The classification evaluation for article prediction generates multi-class classification assessment for the best model which is shown in Table VI.



TABLE V. LOGISTIC REGRESSION AND BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS MODEL TESTING RESULTS

Model Name	AUC	Accuracy	Precision	Recall	F1-Score
<i>Logistic Regression</i>	0.837542	0.839635	0.837542	0.401299	0.8381
<i>BERT</i>	0.958385	0.877076	0.877822	0.877076	0.877262

Note: AUC = area under the receiver operating characteristic curve.

TABLE VI. MODEL EVALUATION OF MULTI-CLASS CLASSIFICATION

Model Name	Class	Precision	Recall	F1 Score	Support
<i>BERT</i>	Negative	0.801749	0.884244	0.840979	311
	Neutral	0.884444	0.866725	0.875495	1148
	Positive	0.902724	0.897485	0.900097	1551
	Accuracy	-	-	0.884385	3010
	Macro avg	0.862972	0.882818	0.872190	3010
	Weighted avg	0.885319	0.884385	0.884606	3010

Uni-Aspect Distribution

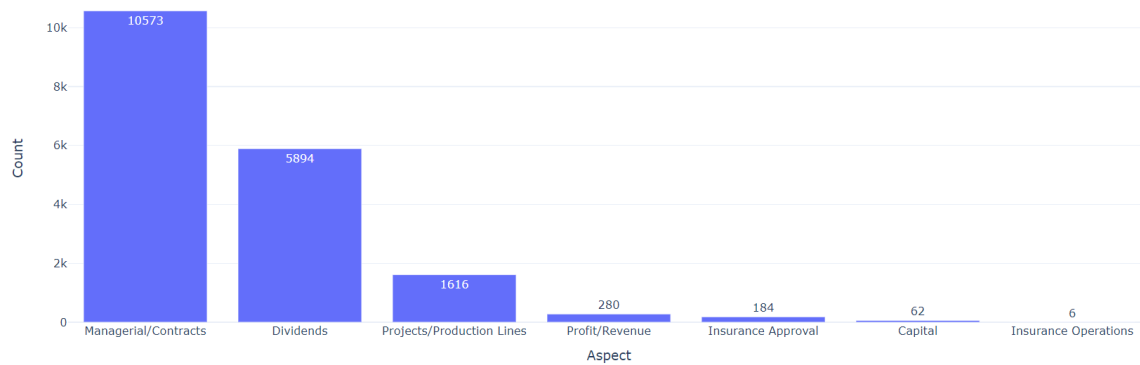


Fig. 2. Highest uni-aspect scoring

Bi-Aspect Distribution

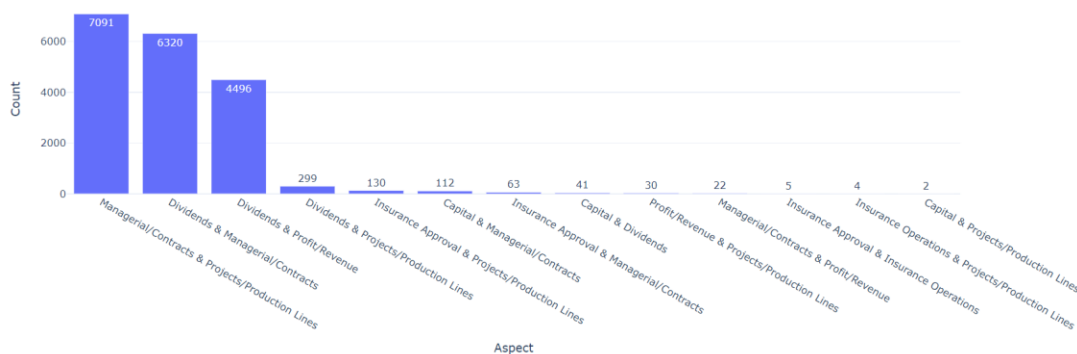


Fig. 3. Highest bi-aspect scoring

### B. Scoring of the Economic Aspects

The results show the percentages of similarity and the highest uni-aspect and bi-aspects for each article's seven aspects. The highest uni-aspect is the aspect that has the highest percentage of similarity. Also, the bi-aspects have the two aspects with the highest percentages of similarity to each article. After the model was applied to the data, the following were the highest uni-aspects among all the aspects: profit (16,799 articles), insurance approval (1,570 articles), capital (195 articles), and insurance operations (51 articles) (see Fig. 2). Moreover, the highest bi-aspects among all the aspects were insurance approval and profit (8,641 articles), and the lowest was insurance approval and operations (1 article) (see Fig. 3). Hence, in each of 2,000 articles, the two highest economic aspects were manually labeled to validate the model's performance. The model was tested on the basis of the labeled data to examine its prediction performance, and its prediction achieved 80% accuracy. The accuracy degree is calculated loosely, such that if the actual class is one of the two predicted classes (bi-aspects), the prediction is accurate. The calculation of accuracy is plausible because the aspects are not mutually exclusive and there is more than one reason in most cases. Therefore, the two highest scores are extracted according to the model, and if the actual aspect is one of the two predicted aspects, the prediction is correct.

C. Polarity Distribution over the Economic Aspect

The final results of both models show the articles' polarity prediction and highest economic aspect as the reason for its negativity or positivity. For instance, the first article in the data is manually labeled positive, predicted by BERT as positive. Its predicted percentages of similarity are highest (79%) for the profit aspect and lowest (63%) for the capital aspect. Also, the highest uni-aspect is the profit aspect, and the highest bi-aspect is profit and production lines.

The insurance approval aspect contains the highest positive news because most of its articles are about the approval of companies' insurance applications, which is good news. Also, the profit aspect has a high percentage of positive articles because most of its articles are about companies' profit and revenue news, which are mostly positive. The insurance operations aspect has the highest percentage of negative news because most of its articles are about the

occurrence of problems and the need for insurance to limit issues, which is bad news. Also, the capital aspect has a high percentage of negative articles because most of its articles are about capital loss news, which are mostly negative (see Fig. 4).

Furthermore, the profit and production lines bi-aspect has the highest percentage of positive news because most of its articles are about companies' profits and revenue from their new projects or production lines, which are good news. The insurance approval and profit bi-aspect also has a high percentage of positive articles because most of its articles are about companies' revenue from the approval of their insurance applications, which are mostly positive. The capital and insurance operations bi-aspect has the highest percentage of negative news because most of its articles are about the occurrence of problems, insurance need, and capital loss, which are mostly negative (see Fig. 5).

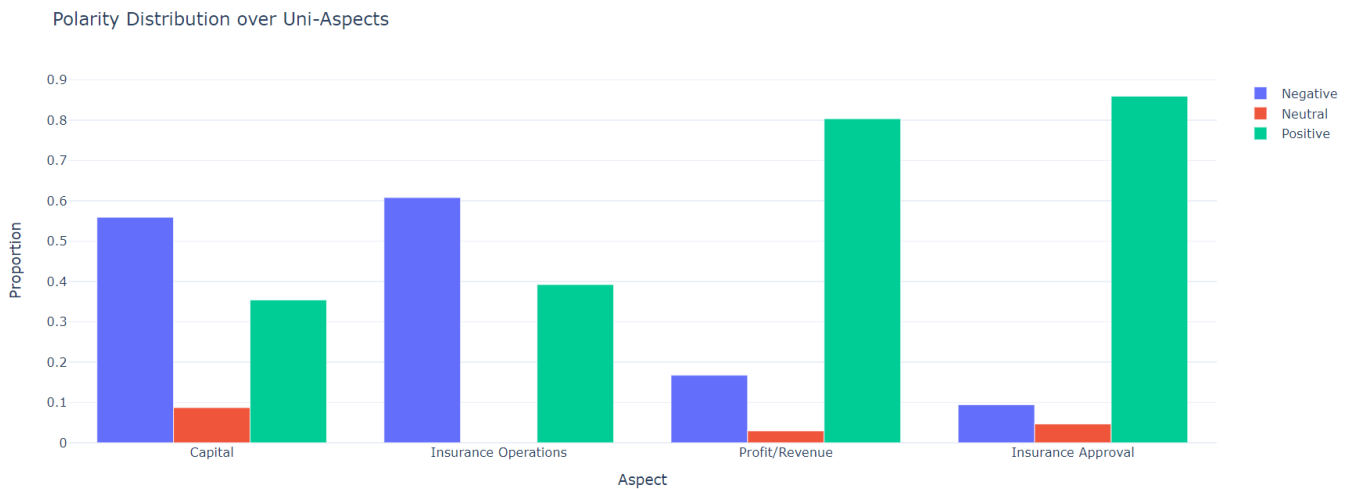


Fig. 4. Uni-aspect polarity distribution.



Fig. 5. Bi-aspect polarity distribution.

## V. CONCLUSION

The present study aimed to predict the polarity classification of stock market news and its economic aspect, which affects the polarity. A couple of models were executed to achieve the study aim: the supervised ML model and the unsupervised mean Word2Vec encoder. These models were applied on a prepared and annotated dataset extracted from the Saudi stock market news platform. The supervised ML model includes the logistic regression and BERT models for classifying the sentiments of stock news articles. The unsupervised model, on the other hand, conducts k-means clustering using the universal ML sentence encoder and Word2Vec. The models were applied to textual Arabic stock news by tuning the hyperparameters and features. The logistic regression model was used with lemmatization, trigrams, and the TFIDF vectorizer for the dataset's features to decrease the token numbers. Such model achieved 84% prediction accuracy by setting the maximum features' value (250,000 tokens). The BERT model, on the other hand, achieved 88% prediction accuracy, the highest outcome in a short time for the classification task. The unsupervised model prefers k-means clustering based on Word2Vec to the universal ML sentence encoder to exhibit semantic unity. Unlike the mean Word2Vec encoder, however, k-means clustering using the universal ML sentence encoder cannot separate the semantics from the sentiments. K-means clustering uses the mean Word2Vec vectorizer to choose ten groups that provide the best semantic separation between aspect categories. The manually extracted economic aspects are profit/revenue, projects/production lines, capital, dividends, insurance operations, insurance approval, and managerial/contracts. Moreover, k-means clustering was tested to predict the aspects of the article on the basis of Word2Vec and the attention algorithm performance. Some examples of the results show that the algorithms tend to prefer the dividends aspect to the other aspects. However, when the algorithm deviates from the correct aspect, it predicts correlated aspects. Finally, the mean Word2Vec encoder achieved 80% economic-aspect prediction accuracy. The developed models are valuable resources that classify the Arabic Stock Market news based on their polarity and main economic aspects. The SA and economic aspects extraction models help to perceive the risks based on the news of the stock. They support making the right decision based on the stocks articles sentiments and their main economic reasons of the polarity. Thus, automated decision-making supports predicting the upcoming stock price trends in investing or analysts' evaluation.

## VI. DATA AVAILABILITY

Data is available in this link based on the method of data available on the request:  
[www.kaggle.com/dataset/bf24521a3898714597a13efa27a85fa208c96ad49620c787c52720861ddd1e6c](https://www.kaggle.com/dataset/bf24521a3898714597a13efa27a85fa208c96ad49620c787c52720861ddd1e6c).

## REFERENCES

[1] Jin, F., Wang, W., Chakraborty, P., Self, N., Chen, F., & Ramakrishnan, N., "Tracking Multiple Social Media for Stock Market Event Prediction," in *Advances in Data Mining. Applications and Theoretical Aspects*, Cham, pp. 16–30, 2017.

[2] Wanjari, Y. W., Mohod, V. D., Gaikwad, D. B., & Deshmukh, S. N., "Automatic news extraction system for Indian online news papers," in

*Proc. - 2014 3rd Int. Conf. Reliab. Infocom Technol. Optim. Trends Futur. Dir. ICRITO 2014*, pp. 1–6, 2015.

[3] A. Mukwazvure and K. P. Supreethi, "A hybrid approach to sentiment analysis of news comments," in *2015 4th Int. Conf. Reliab. Infocom Technol. Optim. Trends Futur. Dir. ICRITO 2015*, pp. 1–6, 2015.

[4] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of Twitter data for predicting stock market movements," in *Int. Conf. Signal Process. Commun. Power Embed. Syst. SCOPES 2016 - Proc.*, pp. 1345–1350, 2017.

[5] F. Hemmatian and M. K. Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," *Artif. Intell. Rev.*, vol. 52, no. 3, pp. 1495–1545, 2019.

[6] Zhang, W., Li, C., Ye, Y., Li, W., & Ngai, E. W., "Dynamic Business Network Analysis for Correlated Stock Price Movement Prediction," *IEEE Intelligent Systems*, vol. 30, no. 2, pp. 26–33, 2015.

[7] D. D. Wu, L. Zheng and D. L. Olson, "A Decision Support Approach for Online Stock Forum Sentiment Analysis," *IEEE transactions on systems, man, and cybernetics: systems*, vol. 44, no. 8, pp. 1077–1087, 2014.

[8] S. Krishnamoorthy, "Sentiment analysis of financial news articles using performance indicators," *Knowl. Inf. Syst.*, vol. 56, no. 2, pp. 373–394, 2018.

[9] P. Choudhari, "Sentiment Analysis and Machine Learning Based Sentiment Classification: A Review," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 3, 2017.

[10] K. Min and H. Moon, "Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network," *IEEE Access*, vol. 6, pp. 55392–55404, 2018.

[11] Al-Ayyoub, M., Nuseir, A., Alsmearat, K., Jararweh, Y., & Gupta, B., "Deep learning for Arabic NLP: A survey," *J. Comput. Sci.*, vol. 26, pp. 522–531, 2018.

[12] Obied, Z., Solyman, A., Ullah, A., Fat'hAlalim, A., & Alsayed, A., "BERT Multilingual and Capsule Network for Arabic Sentiment Analysis," in *Proc. 2020 Int. Conf. Comput. Control. Electr. Electron. Eng. ICCCEEE 2020*, pp. 1-6, 2021.

[13] A. Abuzayed and H. Al-Khalifa, "Sarcasm and Sentiment Detection In Arabic Tweets Using BERT-based Models and Data Augmentation," in *Proc. Sixth Arab. Nat. Lang. Process. Work.*, pp. 312–317, 2021.

[14] M. El-Masri, N. Altrabsheh and H. Mansour, "Successes and challenges of Arabic sentiment analysis research: a literature review," *Soc. Netw. Anal. Min.*, vol. 7, no. 1, pp. 1–22, 2017.

[15] Hmeidi, I., Al-Ayyoub, M., Mahyoub, N. A., & Shehab, M. A., "A lexicon based approach for classifying Arabic multi-labeled text," *Int. J. Web Inf. Syst.*, vol. 1011, no. 17, pp. 324–342, 2016.

[16] J. Kordonis, S. Symeonidis and A. Arampatzis, "Stock Price Forecasting via Sentiment Analysis on Twitter," in *Proc. 20th Pan-Hellenic Conf. Informatics - PCI '16*, pp. 1–6, 2016.

[17] D. de França Costa and N. F. F. da Silva, "INF-UFG at FiQA 2018 Task 1: predicting sentiments and aspects on financial tweets and news headlines," in *Companion Proceedings of the The Web Conference 2018*, pp. 1967–1971, 2018.

[18] L. Qiu, Q. Lei and Z. Zhang, "Advanced Sentiment Classification of Tibetan Microblogs on Smart Campuses Based on Multi-Feature Fusion," *IEEE Access*, vol. 6, pp. 17896–17904, 2018.

[19] Troiano, L., Villa, E. M., & Loia, V., "Replicating a Trading Strategy by Means of LSTM for Financial Industry Applications," *IEEE Trans. Ind. Informatics*, vol. 14, no. 7, pp. 3226–3234, 2018.

[20] Guo, Y., Han, S., Shen, C., Li, Y., Yin, X., & Bai, Y., "An Adaptive SVR for High-Frequency Stock Price Forecasting," *IEEE Access*, vol. 6, pp. 11397–11404, 2018.

[21] P. Pai, S. Member and C. Liu, "Predicting Vehicle Sales by Sentiment Analysis of Twitter Data and Stock Market Values," *IEEE Access*, vol. 6, pp. 57655–57662, 2018.

[22] F. Z. Xing, E. Cambria and R. E. Welsch, "Intelligent asset allocation via market sentiment views," *IEEE Comput. Intell. Mag.*, vol. 13, no. 4, pp. 25–34, 2018.

[23] Y. Touzani, K. Douzi and F. Khoukhi, "Stock Price Forecasting: New Model for Uptrend Detecting and Downtrend Anticipating Based on

- Long Short-Term Memory,” In *Proceedings of the 2018 2nd International Conference on Cloud and Big Data Computing*, pp. 61–65, 2018.
- [24] Piryani, R., Madhavi, D., & Singh, V. K., “Analytical mapping of opinion mining and sentiment analysis research during 2000 – 2015,” *Information Processing & Management*, vol. 53, no. 1. pp. 122-150, 2017.
- [25] Ni, Y. Q., Zhou, H. F., & Ko, J. M., “Generalization Capability of Neural Network Models for Temperature-Frequency Correlation Using Monitoring Data,” *J. Struct. Eng.*, vol. 135, no. 10, pp. 1290-1300, 2009.
- [26] M. Anthony, and P. L. Bartlett, "Neural network learning: Theoretical foundations," in *Cambridge: cambridge university press*, vol. 9. 1999.
- [27] A. Yadav and D. K. Vishwakarma, “Sentiment analysis using deep learning architectures: a review,” *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 4335–4385, 2020.
- [28] Sachin, S., Tripathi, A., Mahajan, N., Aggarwal, S., & Nagrath, P., “Sentiment Analysis Using Gated Recurrent Neural Networks,” *SN Comput. Sci.*, vol. 1, no. 2, pp. 1–13, 2020.
- [29] J. V. Tembhurne and T. Diwan, “Sentiment analysis in textual, visual and multimodal inputs using recurrent neural networks,” *Multimed. Tools Appl.*, vol. 80, no. 5, pp. 6871–6910, 2021.
- [30] Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., & Mosavi, A., “Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; A Comparative Analysis,” *IEEE Access*, vol. 8, pp. 150199–150212, 2020.
- [31] Cai, R., Qin, B., Chen, Y., Zhang, L., Yang, R., Chen, S., & Wang, W., “Sentiment analysis about investors and consumers in energy market based on BERT-BILSTM,” *IEEE Access*, vol. 8, pp. 171408–171415, 2020.
- [32] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I., “Attention is all you need,” *Adv. Neural Inf. Process. Syst.*, pp. 5999–6009, 2017.
- [33] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K., “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, pp. 4171–4186, 2019.
- [34] W. M. Szu, Y. C. Wang and W. R. Yang, “How does investor sentiment affect implied risk-neutral distributions of call and put options?,” In *HANDBOOK OF FINANCIAL ECONOMETRICS, MATHEMATICS, STATISTICS, AND MACHINE LEARNING*, vol. 18, no. 2, pp. 1599-1636, 2015.
- [35] M. Koppel and J. Schler, “The importance of neutral examples for learning sentiment,” *Comput. Intell.*, vol. 22, no. 2, pp. 100–109, 2006.
- [36] T. S. Ng, “Machine learning,” *Stud. Syst. Decis. Control*, vol. 65, pp. 121–151, 2016.
- [37] G. Hackeling, "Mastering Machine Learning with scikit-learn," in *Packt Publishing Ltd*, 2017.
- [38] P. Liang and M. I. Jordan, “An asymptotic analysis of generative, discriminative, and pseudolikelihood estimators,” in *Proc. 25th Int. Conf. Mach. Learn.*, pp. 584–591, 2008.
- [39] F. A. Gers, J. Schmidhuber and F. Cummins, “Learning to forget: Continual prediction with LSTM,” *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000.
- [40] I. Goodfellow, Y. Bengio and A. Courville, “deep learning English version,” MIT press, p. 800, 2017.

# e-Government Usability Evaluation: A Comparison between Algeria and the UK

Mohamed Benaida

Faculty of Computer and Information Systems  
Islamic University of Madinah, Saudi Arabia

**Abstract**—e-Government holds the keys to improving government services provided to citizens and the private sectors within their countries. Although Algeria is the largest country in Africa and has one of the most thriving economies in the continent, it is remarkable that the Algerian EGDI ranking was 120th according to the latest UN e-government survey. This inspired the researcher to investigate the relationship between the success factors of e-services in developed countries and their counterparts in developing countries. The main aim of this study is to explore the factors that influence the level of usability of e-government services between developing and developed countries against a set of specific guidelines to provide means for improving these services in developing countries. The researcher selectively extracted three guideline categories from Research-Based Web Design and Usability Guidelines as a means for expert evaluation of 10 Algerian e-government services compared to British e-government services. Our results show that Algerian e-services lack mostly in Use Frames when Functions Must Remain Accessible, Highlighting Information, and Graphics Should Not Look like Banner Ads (belonging to Page Layout, Text Appearance, and Graphics, Images & Multimedia respectively), whereas UK e-services scored highly across all three categories. These findings further enhance the UN e-government survey and identifies the sub-categories that developing countries need to pay more attention to in order to provide a more reliable and robust e-service to its users and citizens. Furthermore, this study proposes that the Research-Based Web Design & Usability Guidelines can be converted into an evaluation tool to be used by evaluators to easily assess the usability of a website. The combination of relative importance, chapters of the guidelines, and their respective guidelines gathered from Research-Based Web Design & Usability Guidelines, along with the evaluation of these individual guidelines by evaluators will serve as an integral tool for developers when developing e-government services to reach the satisfaction of the users.

**Keywords**— *Human computer interaction; usability evaluation; web design; e-Government; user satisfaction*

## I. INTRODUCTION

The internet officially launched in Algeria in 2001, Seven years later in 2008, e-government services in Algeria were established [1]. Only recently has the Algerian government tried to keep up with modern technology by using online services. However, the shift from centralisation to using electronic services has not been an easy process. Design, logistics infrastructure, internet speed, and adoption of new services; are all factors that are considered great challenges facing the government as it moves into the world of electronic services. Like other governments in developing countries, the

Algerian government seeks to establish an electronic system through which citizens can benefit from the e-services provided by their government in various fields. However, the e-government project encountered many problems regarding infrastructure, use of technology in a sound manner, the quality of services, and ease of access for citizens. The ultimate aim of e-governments is to improve the usability of their services for their users, thus increasing user interaction. There are several definitions of e-government based on its use, however this research focuses on e-government as a system rather than from a technical or social aspect. One definition describes e-government as "the use of information technology to enable and improve the efficiency of government-provided services to citizens, employees, businesses and agencies" [2]. Most governments in developing countries have recently engaged in e-services that aim to conserve effort, time, and money for the population. Many various usability guidelines exist today (such as ISO/CD 9241-151). Thus, it is hard to point out one guideline as the model since every audience around the world has its own culture, background, symbols, language, and political system [3]. Furthermore, the level of the country's economy affects their perception and judgments of everything around them, and therefore each audience has different needs [4], [5].

This study decided to utilise the Research-Based Web Design & Usability Guidelines (RBED&UG) due to its exclusivity and popularity in government agencies, educational programs, and various independent sectors. These guidelines were established by the U.S. Department of Health and Human Services initially in 2003 and further developed in 2006 with the aim of helping designers develop a large range of websites, but mainly information filled government services. The exclusivity of these guidelines lies in their relative importance in rankings whilst also providing references and expert feedback to support their argument. Another reason this study decided to use guidelines is due to the recommendations provided by the authors of these guidelines, which motivates researchers to explore avenues with little or almost no research, as is the case is with Algerian e-government. The latest edition is made up of 18 main chapters, which are divided into multiple guidelines enhanced by the relative importance and strength of evidence providing a total of 209 guidelines across these chapters [6]. This research focused on three chapters of this guideline (Page Layout, Text Appearance, and Graphics, Multimedia, & Images) as it aimed to provide accurate and in-depth results rather than a broad and general set of results which may not be as useful to the web designers of these websites.

The main reasons for picking Algerian e-government services as the subject of this study is that there is a lack of research in this field regarding less developed countries generally [7], and according to our knowledge, even after more than a decade since its establishment, there are no studies that analyse the usability of e-government services in Algeria. Furthermore, this simultaneously fulfils the recommendations of the authors of the RBED&UG of researching into new areas and providing relevant research results. Therefore, the main contribution of this research is that it is the first study to shed light on the usability of Algerian e-government services and provide web designers with a deep insight into how to address the flaws (lack of online services, lack of visual consistency, etc.) which the experts in this study have identified in order to help the government create usable, learnable, and effective e-services that will increase user interaction.

The e-government development index (EGDI) measures a country's capability to deliver online public services and is calculated using three variables including telecommunications infrastructure index (TII), human capital index (HCI), and online service index (OSI). This study compares the Algerian government e-services with the government e-services in the United Kingdom, as Britain is a developed country and ranks high in the EGDI rankings published in the latest United Nations e-government survey so that a comparison can be made between Algeria as a developing country and Britain as a developed country, allowing for better understanding of deficiencies and focusing on the factors that are likely to have an impact on improving the quality of service provided to citizens in developing countries, which facilitates the lives of citizens and makes them more efficient and of better quality. The British are 7th in the EGDI rankings whereas Algeria is ranked 120th in the world [8]. The aim is to explore why there are such big gaps in the e-services provided between developed and less developed countries. Unlike the British e-government services which are all available on one website, Algerian e-government services are all independent and each website serves a different purpose, for example, the ministry of interior, the ministry of sports etc. Therefore, this study selected ten Algerian e-government websites to compare with the British e-government website to serve as a realistic representation of Algerian e-services. The null hypothesis in this study is that there is no significant difference between Algerian and British e-government services, whereas the alternative hypothesis states that there is a significant difference between Algerian and British e-government services.

Another contribution of this study is that it converts the RBED&UG into a questionnaire to be used by web designers to help evaluate and gather relevant data for improving electronic web services. Researchers realised the importance of these guidelines in developing and initiating web systems in various fields and the high potential that it can be used by web designers for evaluating the level of a system and addressing certain issues that may present. Researchers believe that the same concept of converting guidelines into a questionnaire to be used as an evaluating tool for web designers can be implemented with any of the chapters of the RBED&UG. The study only used three chapters as a representative of all

chapters and to show that it can indeed be implemented efficiently and effectively to good use.

This research paper is split into six parts including the introduction which provides an overview of the main components of this research. Then, the literature review dwells on previous research and identifies their key findings and shortcomings which need to be addressed. The methodology details the stages of this research and the demographic data of the participants that were involved. Results and discussion include a detailed analysis which breaks down the various attributes that are suffering or excelling in the usability of e-government services. The conclusion provides an overview of our research along with the main results and contributions it brings forward to the domain of e-government usability. Finally, the limitations and future work marks the end of this research paper.

## II. LITERATURE REVIEW

Information technology is growing rapidly; government services must also be up to date with these technological advancements. In developing countries, this challenge is far greater, since many still suffer from poor IT infrastructure and financial problems, which hinder their progress in providing advanced electronic services to their citizens. Besides, building trust between the citizens and using electronic services is still in development [9]. This section discusses e-government websites and usability, usability evaluation methods and e-governments Algerian websites.

### A. e-Government Websites and Usability

In the last decade, e-government services have become more important than ever before. Many governments have launched at least one or more systems that provide citizens with the necessary information and adequate services. It summarises and stores relevant information based on the quality of the internet and the level of ability of the users to use their internet services. Many factors can affect the performance of these e-government systems such as the educational background of the users, culture, and simplicity of the system, efficiency, robustness and reliability [3]. There are various definitions of e-government depending on the field of interest and the perception of experts [10]. The World Bank (2015) believes that "E-Government refers to the use by government agencies of information technologies (such as Wide Area Networks, the Internet, and mobile computing) that have the ability to transform relations with citizens, businesses, and other arms of government" [11]. However, a comprehensive study on multiple definitions of e-government established six main elements that serve as a widely shared definition. These variables consist of; "(1) the major initiatives of management and delivery of information and public services. (2) Taken by all levels of governments (including agencies and sectors). (3) On behalf of citizens, business. (4) Involving using multi-ways of internet, web site, system integration, and interoperability. (5) To enhance the services (information, communication, policy making), quality and security. (6) As a new key (main, important) strategy or approach" [10].

Usability is defined by the ISO as the "extent to which a system, product or service can be used by specified users to

achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [12]. Any new system or product should achieve its goals and satisfy its end users by ensuring that the system is more interactive and attractive to maximise the benefits of the system. Most e-government projects (35% total failures and 50% partial failures) in less developed countries are unsuccessful [7]. They are unable to fulfil the satisfaction of their users due to the widespread usability issues found on their systems [13]. Nielsen’s description of usability via the implementation of various metrics is widely accepted, thus ten usability guidelines were initiated to identify errors, evaluate the level of effectiveness and efficiency and establish user control, thus, ultimately fulfilling user satisfaction.[3].

Some studies [14] argue that e-government can benefit from usability in two aspects; satisfaction is key due to improved user performance on e-government services, on the other hand, the first impressions are also crucial for attracting users to using e-government online services. e-Government relates to usability in multiple ways correlating to satisfaction; internet speeds [15] loading speed [16] usefulness and ease of use [17]. The abovementioned variables are one of the most influential in the usability of e-government according to previous studies. Poor design interface can affect not only the perception of users, but in some cases can become more serious and influence various factors such as user safety [18]. Therefore, users should be an essential part in the design stages to minimise the errors that end users can face [7] explore the value of e-government and the public value it holds. Through this research, six values were identified; Open Government capabilities, improved ethical behaviour and professionalism, improved public services, improved administrative efficiency, improved social value and well-being, and improved trust and confidence in government. In least developed countries and developing countries, there was a lack of research in attempting to figure out the public value of e-government and the user perception of these services.

Many studies in other countries have attempted to figure out solutions and identify factors that can influence the quality and usability of e-governments services. Studies have shown that usability is the heart of an e-government system that attempts to fulfil the satisfaction of its users. Furthermore, it can directly affect the trust of the users [19]. The meaning of usability may not be clear enough. However, usability is not just how to make the system easy to use; instead, it is the ease of use based on efficiency, effectiveness, learnability, and user satisfaction [20]. Some studies argue that western e-government websites fail to deliver services to their citizens [21]. There were several usability issues identified in developing countries such as Jordan [22], Malaysia [23], China [24], and Sub-Saharan Africa [25] such as South Africa [26] and Tanzania [27]. The usability problems identified included accessibility [28] poor design [29] learnability [30] and text appearance [31]. Furthermore, poor usability design affects the quality of the website and the adoption of users, which drives them away from revisiting the website [5].

Each country faces its own specific usability issues, and many studies have found different factors that can positively or negatively evaluate the level of e-government services and the

quality provided to their citizens. e-Government services in Mauritius Island showed that factors like expected performance and facilitating conditions have a positive relationship with the users' behavioural intentions. On the other hand, computer self-efficacy is negatively related to the users' behavioural intentions. The authors suggest essential steps the government could take to help the citizens adopt their e-government services effectively and efficiently [32]. In Turkey, factors that could help citizens adopt e-government services are Performance Expectancy, Social Influence, Facilitating Conditions, and Trust on the Internet. Authors [33] have initiated a new guideline for policymakers to initiate attractive e-government services by considering the highlighted priorities of the citizens. The authors proposed a new analysis technique that could potentially increase and maximise e-government portals' potential to benefit individuals from various cultural backgrounds. This analysis technique incorporates various factors such as content analysis, user perception, and persuasive quality gap. Culture is another factor that can affect the usability satisfaction level [34], [16]. For example, [35] argue that their suggested approach provides an estimation method to measure the contribution of e-government quality attribute to cross-cultural quality gap. They also examine the satisfaction level of Jordanian citizens when using these e-government services. Several factors influencing satisfaction were identified; security and privacy, trust, accessibility, awareness, and quality. They claim that their findings have proven very useful for industry practitioners and policymakers within the government.

#### *B. Usability Evaluation Methods*

Usability evaluation becomes more crucial in testing usability issues including content and page layout, to create a system that suits the requirements of the user and meets their needs. There are two main usability evaluation methods used to assess usability websites: heuristic evaluation and usability testing.

Many studies use this type of evaluation (see Table I) via Nielsen’s ten heuristics due to its low cost, time efficiency, and how it requires only a few experts to give their opinion about the problems that face certain applications or websites. On the other hand, “usability testing refers to evaluating a product or service by testing it with representative users” [36]. Usability testing gathers a group of users to complete several tasks on the system to assess the usability level and identify any issues within it. The users assessing the system are the same users that the designer is creating a usable system for, thus any issues identified by these participants are crucial and provide a real-life representation of the end user feedback after the official launch of the system. Consequently, designers can take their feedback on board and make the appropriate adjustments to improve the overall system to make it more usable and attractive. For these reasons, this study proposes that the new questionnaire is to be used as a heuristic evaluation method used by evaluators in web design.

TABLE I. PREVIOUS STUDIES OF USABILITY EVALUATION METHODS

Study	Method	Type
Fonseca and Peñalvo (2019) [37]	Usability testing	Gaming
Huang and Chen (2019) [38]	Usability testing	Digital learning environments
Fuller-Tyszkiewicz et al (2018) [39]	Usability testing	Health app
Krzewińska et al (2018) [40]	Usability testing	Web systems
De Souza Filho et al (2017) [41]	Collaborative heuristic evaluation	Gaming
Diaz et al (2017) [42]	Heuristic Evaluation	e-Commerce
Jucá et al (2017) [43]	Heuristic Evaluation	Gaming
Alhadreti and Mayhew (2017) [44]	Usability testing	Examine think-aloud methods
Murillo et al (2017) [45]	Usability testing and Heuristic Evaluation	Web-based systems
Atashi et al (2016) [46]	Heuristic Evaluation	User Interface
Falkowska et al (2016) [47]	Usability testing (Eye Tracking)	Web applications
Inostroza et al (2015) [48]	Heuristic Evaluation	Smartphones and application
Chynał and Sobecki (2015) [49]	Usability testing	User Interface
Bezerra et al (2014) [50]	Usability testing	Ubiquitous systems
Ko et al (2013) [51]	Heuristic Evaluation	Mobile app
Boothe et al (2013) [52]	Usability testing	Interface medium
Sivaji et al (2011) [53]	Heuristic Evaluation	e-Government websites
Thyvalikakath (2009) [54]	Usability testing and Heuristic Evaluation	Comparison study
Garcia et al (2005) [55]	Heuristic Evaluation	e-Government

### III. METHODOLOGY

#### A. e-Services Evaluated

Ten Algerian e-Government websites were selected for evaluation in this study. They include the following: Minister of Defence, Minister of Interior and Local Government, Minister of Foreign Affairs, Minister of Finance, Minister of Energy, Minister of War Veterans (Moudjahidine), Minister of Religious Affairs and Endowments (Wakfs), Minister of Vocational Education and Training Professionals, Ministry of Culture and Arts, and Minister of National Solidarity, Family and Women's Affairs. There is no unified website in Algeria that can provide more government electronic services, so the researcher had to collect the ten most used Algerian websites to evaluate electronic government services, while in the United Kingdom there is only one website that collects government

electronic services, and accordingly, a comparison was made between these ten sites. After this information was evaluated by experts, the average was collected from each of the ten sites and compared with the standard British site in order to facilitate the evaluation process.

#### B. Proposed Usability Questionnaire Model

After realising the effectiveness of the RBED&UG, researchers believed that these guidelines could be effectively used during usability of e-government services to identify flaws within the usability of these systems. Whilst guidelines are used by web designers when creating a web page, researchers propose that the RBED&UG is transformed and used as a questionnaire to evaluate the system after it is created so that web designers can brush up and improve on the areas that are lacking thus providing a more robust and reliable service. This study decided to selectively examine the effectiveness of three chapters (Page Layout, Text Appearance, Graphics, Images & Multimedia) from the RBED&UG to serve as a strong representative for the other 15 chapters. Since the original guidelines follow a five-point scale, researchers believed it would be most appropriate to be consistent and implement a five-point scale in the questionnaire where one represents "very poor" to five which represents "very strong". Tables II, III and IV display the questionnaires that were provided to the experts before the start of the experiment. They were provided with two copies of each one, totalling to six copies of these questionnaires to evaluate both Algerian and British e-Government services relative to the three selected guidelines. The experts marked their selected scores for the relative e-service and guidelines of these categories. These results will be collated by the researcher at the end of their evaluation and utilised for statistical analysis of the data.

TABLE II. TABLE TYPE STYLES

Usability Page Layout Questionnaire					
	1	2	3	4	5
1. Avoid Cluttered Displays					
2. Place Important Items Consistently					
3. Place Important Items at Top Centre					
4. Structure for Easy Comparison					
5. Establish Level of Importance					
6. Optimize Display Density					
7. Align Items on a Page					
8. Use Fluid Layouts					
9. Avoid Scroll Stoppers					
10. Set Appropriate Page Lengths					
11. Use Moderate White Space					
12. Choose Appropriate Line Lengths					
13. Use Frames when Functions Must Remain Accessible					



TABLE III. USABILITY TEXT APPEARANCE QUESTIONNAIRE

Usability Text Appearance Questionnaire					
	1	2	3	4	5
1. Use Black Text on Plain, High-Contrast Backgrounds					
2. Format Common Items Consistently					
3. Use Mixed-Case for Prose Text					
4. Ensure Visual Consistency					
5. Use Bold Text Sparingly					
6. Use Attention-Attracting Features when Appropriate					
7. Use Familiar Fonts					
8. Use at Least 12-Point Font					
9. Color-Coding and Instructions					
10. Emphasize Importance					
11. Highlighting Information					

TABLE IV. USABILITY GRAPHICS, IMAGES, & MULTIMEDIA QUESTIONNAIRE

Usability Graphics, Images, & Multimedia Questionnaire					
	1	2	3	4	5
1. Use Simple Background Images					
2. Label Clickable Images					
3. Ensure that Images Do Not Slow Downloads					
4. Use Video, Animation, and Audio Meaningfully					
5. Include Logos					
6. Graphics Should Not Look like Banner Ads					
7. Limit Large Images Above the Fold					
8. Ensure Web Site Images Convey Intended Messages					
9. Limit the Use of Images					
10. Include Actual Data with Data Graphics					
11. Display Monitoring Information Graphically					

12. Introduce Animation					
13. Emulate Real-World Objects					
14. Use Thumbnail Images to Preview Larger Images					
15. Use Images to Facilitate Learning					
16. Using Photographs of People					

### C. Procedure

The volunteering experts were provided with all the websites required for this experiment and a brief was given before the beginning of the evaluation process along with copies of the questionnaires they would have to fill out. The brief very simply explained the goals of our research and that the participants are testing the guidelines as a questionnaire for evaluation to explore the issues in Algerian and British e-Government websites. The results would also simultaneously help assess the validity of the proposed questionnaire. The researcher decided to allocate some time for the experts before evaluation to get familiar with these e-services so that they can easily and effectively evaluate them once the process begins. After that the experts evaluated the e-services according to the three dimensions (Page Layout, Text Appearance, and Graphics, Images, & Multimedia) from our usability questionnaire accordingly, freely with no time constraints. Experts were asked to return the questionnaire once complete.

### D. Participants

Since some e-services are only available in English, Arabic, or French, the researcher decided to select seven experts who are confident with all three languages to help gather the results of our study across all platforms with no issues and full understanding. Each expert evaluated the ten selected Algerian e-Government websites and the British e-Government website. The participants who contributed to this research voluntarily are from different educational levels and gender backgrounds. Most participating experts were males aged between 36 and 50 years who are of a postgraduate degree status. Table V shows the expert demographics.

TABLE V. EXPERT DEMOGRAPHICS

Demographic	Category	Number of Participants (7)	%
Age	25-35	1	14
	36-50	4	57
	50+	2	29
Gender	Women	2	29
	Men	5	71
Education Level	Undergraduate Degree	2	29
	Postgraduate Degree	5	71

### E. SPSS Analysis

The results of the expert evaluation gathered via the proposed questionnaire were inserted into the SPSS program and results of importance to this study were extracted such as the paired samples statistics mean and standard deviation, the paired samples test mean and standard deviation, t-values, and p-values. The paired samples statistics mean shows us the average score for each guideline of each chapter, whereas the paired samples statistics standard deviation uses the results of the experts to identify the level of consensus between the set of results. The paired samples test mean is the difference between the paired samples statistics mean of the two tested variables. Furthermore, the p values show us the level of significant differences between our tested samples in each element of the three categories selected. For the data to be considered significantly different, the value of p must be less than 0.05, thus rejecting the null hypothesis which in this study is that there is no difference between the Algerian and British e-Government services, therefore the alternative hypothesis would be accepted instead. However, if the value of p is greater than 0.05, then the data is deemed to be insignificantly different, and the null hypothesis is accepted. This analytical technique helps researchers identify where the main flaws lie in a certain set of data, and in this case, it shows us which guideline suffers the most in the usability of Algerian and British e-Services and where the biggest difference lies between the two.

## IV. RESULTS AND DISCUSSION

Experts evaluated Algerian and British e-Government services via our proposed questionnaire derived from the three categories of the RBED&UG. The results of each of the seven experts were gathered and inserted into the statistical software SPSS. This software provided us with some important data which will be analysed later on. As expected from the results, the British e-Government services scored higher across all three categories overall as seen in Fig. 1. The Algerian e-Government services scored 2.748, in contrast to the British e-Government services which scored 4.758 in the Page Layout category. For Text Appearance, the scores for the Algerian and British e-services were 2.785 and 4.562 respectively. Finally, in the Graphics, Images & Multimedia, the results followed a similar pattern, where the Algerian e-Government services were inferior to the British services (2.594 and 4.143 respectively), however this category suffered the most compared to the other two categories. Furthermore, each category will be discussed independently later on, and an in-depth analysis will be provided. The null hypothesis and alternative hypothesis will be accepted or rejected relevantly in each of the three categories.

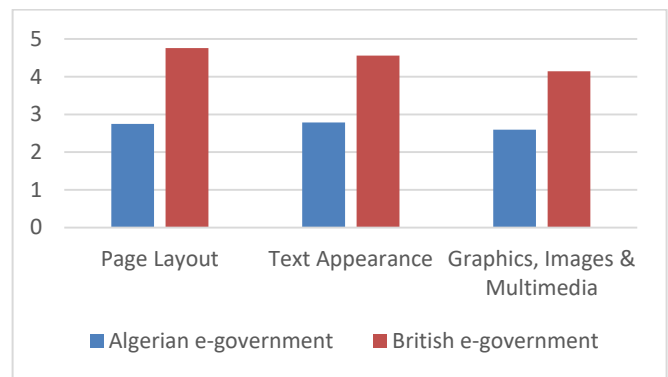


Fig. 1. Chart comparing the mean results of Algerian and British e-Government services across the three selected categories of the RBED&UG

### A. Page Layout

Page layout is the main factor in the user interface; it outlines the position of elements and their dimensions [56]. The most important aspect when designing a page layout is to include necessary information only and to avoid overwhelming the user with unnecessary information that will only lead to confusion and over complication.

With the help of the SPSS software, the researcher was able to extract relevant information as seen in Table VI which presents all this data along with the 13 guidelines of Page Layout that are ranked according to relative importance according to the RBED&UG. The results show that all the guidelines of the Page Layout when comparing the Algerian and British services show significant differences due to the value of p being less than 0.05. Therefore, the British e-Government services according to the experts in this study are significantly different to the Algerian e-Government services. As a result, the null hypothesis for this study is rejected, and the alternative hypothesis is accepted. These results compliment the research conducted by the UN which displays the ranking difference regarding the EGDI, where Algeria falls behind the UK in every attribute of the page layout. Algeria must focus on the page layout as a whole, according to the relative importance of its guidelines to make sure that the most impactful factors of web usability are approached first. According to the experts, the two strongest attributes in the Algerian e-Government services include Establish Level of Importance and Choose Appropriate Line Lengths which both scored 3.571. On the other hand, the experts believed that Choose Appropriate Line Lengths (paired samples statistics mean = 4.929) was the strongest attribute of the British e-Government services. Furthermore, the two lowest scored variables in this section for the Algerian e-Government services include Use Frames when Functions Must Remain

Accessible and Use Moderate White Space which scored 2.429 and 2.500 respectively. Whereas, for the British e-Services, the lowest scored guidelines include Use Frames when Functions Must Remain Accessible and Establish Level of Importance (3.356 and 4.429, respectively). The two highest scoring categories of paired samples test mean values are Structure for Easy Comparison and Use Moderate White Space (both scored -2.143). This variable represents the difference in the level of guidelines between the Algerian and British e-services as identified by the experts. Therefore, this result shows that the biggest difference between the two tested subjects in this study lie in the two abovementioned sections of Page Layout. These results display the areas that require the most attention whilst also simultaneously adhering to the level of importance of these factors to prioritise the level of importance needed to identify which attributes must be first approached or approached with the most attention.

TABLE VI. RESULTS OF THE SPSS ANALYSIS FOR PAGE LAYOUT

Page Layout				
Guidelines	Relative Importance	Paired Samples Statistics Mean (SD) Algerian e-Government & British e-Government	Paired Samples Test Mean (SD)	Significance
1. Avoid Cluttered Displays	5	2.714 (0.184)	-1.929 (0.787)	t=-6.485 p=0.001
		4.642 (0.143)		
2. Place Important Items Consistently	5	3.071 (0.202)	-1.571 (0.731)	t=-5.680 p=0.001
		4.642 (0.180)		
3. Place Important Items at Top Centre	5	2.790 (0.184)	-1.786 (0.756)	t=-6.250 p=0.001
		4.571 (0.170)		
4. Structure for Easy Comparison	4	2.571 (0.202)	-2.143 (0.556)	t=-10.190 p=0.000
		4.715 (0.184)		
5. Establish Level of Importance	4	3.571 (0.170)	-0.857 (0.610)	t=-3.286 p=0.017
		4.429 (0.170)		
6. Optimize Display Density	4	3.286 (0.214)	-1.000 (0.764)	t=-3.464 p=0.013
		4.786 (0.149)		
7. Align Items on a Page	4	2.642(0.180)	-2.143 (0.627)	t=-9.045 p=0.000
		4.786(0.149)		
8. Use Fluid Layouts	3	2.714 (0.240)	-2.000 (0.707)	t=-7.483 p=0.000
		4.714 (0.149)		
9. Avoid Scroll	3	2.857 (0.922)	-1.929	t=-27.000

Stoppers		4.786 (0.149)	(0.189)	p=0.000
10. Set Appropriate Page Lengths	3	2.929 (0.202)	-2.000 (0.646)	t=-8.198 p=0.000
		4.929 (0.714)		
11. Use Moderate White Space	3	2.500 (0.110)	-2.143 (0.378)	t=-15.000 p=0.000
		4.643 (0.180)		
12. Choose Appropriate Line Lengths	2	3.571 (0.202)	-1.357 (0.690)	t=-5.203 p=0.002
		4.929 (0.714)		
13. Use Frames when Functions Must Remain Accessible	1	2.429 (0.130)	-0.929 (0.838)	t=-2.931 p=0.026
		3.356 (0.340)		

### B. Text Appearance

Text is the backbone of any website as it expresses and infers a lot about the website whilst also playing a major role in user satisfaction [57]. Therefore, designers must pay more attention to the main points that affect web design usability [58]. Designers must also research the location of their target audience and study their culture as each culture's needs differ [3]. Therefore, further investigation into the level of infrastructures and the level of user adoption must be undergone. Consequently, creating a website with the appropriate font, font size etc. will allow users to become more comfortable while effectively surfing the website. Consistency, and the level of the quality attributes of the content, which include understandability, relevance, accuracy and finally coverage, should also be focused on accordingly as it is crucial when evaluating the usability of text appearance [59], [60].

Table VII which presents the results of the SPSS analysis for Text Appearance also rejects the null hypothesis of this research as none of the p values of its relative guidelines appeared to be greater than 0.05. The analysis also shows that the Algerian e-services matched the British e-services in one category (Emphasis Importance) which holds a relative importance of 2 according to the RBED&UG. The paired samples statistics mean simply represents the average score of all seven experts in their assessment of the level of the attributes within the websites. The results show that for the Algerian e-Government services, use at Least 12-Point Font (4.357) and Use Similar Fonts (4.071) scored the highest. Whereas, for the British e-Government services, the experts believed that Ensure Visual Consistency and Highlighting Information were the strongest attributes of this category, as all experts unanimously agreed to their perfect score of five. On the contrary, the lowest scoring factors of Text Appearance regarding the Algerian e-services belonged to Highlighting Information (1.856) and Use Attention-Attracting Features when Appropriate (2.071). The experts believed that for the British e-Services, the weakest score belonged to Emphasize

Importance (2.786). In addition, the two highest scoring guidelines regarding the paired samples test mean are Highlighting Information and Ensure Visual Consistency, where they scored -3.143 and -2.857, respectively.

TABLE VII. RESULTS OF THE SPSS ANALYSIS FOR TEXT APPEARANCE

Text Appearance				
Guidelines	Relative Importance	Paired Samples Statistics Mean (SD) Algerian e-Government & British e-Government	Paired Samples Test Mean (SD)	Significance
1. Use Black Text on Plain, High-Contrast Backgrounds	4	2.142 (0.210)	-2.143 (0.801)	t= -7.071 p=0.000
		4.186 (0.149)		
2. Format Common Items Consistently	4	2.714 (0.240)	-2.143 (0.748)	t=-7.579 p=0.000
		4.857 (0.092)		
3. Use Mixed-Case for Prose Text	4	3.714 (0.101)	-0.714 (0.636)	t=-2.970 p=0.000
		4.429 (0.170)		
4. Ensure Visual Consistency	4	2.143 (0.210)	-2.857 (0.556)	t= -13.587 p=0.000
		5.000 (0.000)		
5. Use Bold Text Sparingly	3	2.286 (0.184)	-2.714 (0.488)	t=-14.717 p=0.003
		5.000 (0.000)		
6. Use Attention-Attracting Features when Appropriate	3	2.071 (0.202)	-2.429 (0.450)	t=-14.283 p=0.000
		4.500 (0.154)		
7. Use Familiar Fonts	3	4.071 (0.130)	-0.571 (0.535)	t=-2.828 p=0.000
		4.643 (0.143)		
8. Use at Least 12-Point Font	3	4.357 (0.922)	-0.429 (0.345)	t=-3.286 p=0.000
		4.786 (0.101)		
9. Color-Coding and Instructions	2	2.500 (0.154)	-2.500 (0.408)	t=-16.202 p=0.000
		5.000 (0.000)		
10. Emphasize Importance	2	2.786 (0.184)	-1.857 (0.748)	t=-6.569 p=0.011
		2.786 (0.184)		
11. Highlighting Information	2	1.856 (0.210)	-3.143 (0.556)	t=-14.946 p=0.000
		5.000 (0.000)		

### C. Graphics, Images, and Multimedia

Graphics, Images & Multimedia was the densest category amongst the three tested in this study, as it contained a total of 16 guidelines. Graphics, images, and multimedia individually or collectively, play a major role in user interface design. The proper use of graphics can enhance the usability of websites [61]. Studies show that images are a crucial aspect when designing a website [5], [62]. Images also influence the interactive systems of enhancing the positive or negative reaction of the users of websites [63]. On the other hand, multimedia can attract more users and efficiently summarize the message that the designers and the website owners want to deliver to the users [64]. The designer should know the background of the target users to create a website that respects their culture and values [5]. Furthermore, the designer should look to get familiar with the internet speed of the target region, which will dictate the use or lack of graphics, images and multimedia as they can take a while to upload/download or may not upload/download if the internet speed is not enough or if the file is too big. Therefore, designers must ensure that all these factors are considered to avoid failing their target users and avoid providing them with an unattractive website that their internet connection speed is too slow to handle that may potentially disrespect their culture, which ultimately causes the whole website to fail. The abovementioned factors are crucial to the success of any website.

The results from this category (Table VIII) follow a similar pattern to the previous categories, where according to the experts; the British e-Services are a lot superior to the Algerian e-services. Across all guidelines, there is a significant difference between the two tested subjects. However, for Introduce Animation, the value of p is 0.726 which is greater than 0.05 which means that there is no significant difference between the Algerian and British e-Government services in relation to this guideline. Thus, for this category alone which has a relative importance of 2 according to the RBED&UG, the null hypothesis is accepted, and the alternative hypothesis is rejected. The strongest attributes according to the experts regarding the Algerian e-services are Using Photographs of People (4.429) and Include Logos (4.071), whereas the weakest attributes include Graphics Should Not Look like Banner Ads and Ensure Web Site Images Convey Intended Messages which both scored 1.643 according to the average rating of the experts. The Graphics Should Not Look like Banners guideline had the highest paired samples test mean (-3.357) amongst all guidelines of this category. Overall, the results of this category illustrate the shortcomings of the Algerian e-services and how they should approach each category according to its evaluated level by the experts and the relative importance determined by the guidelines.

TABLE VIII. RESULTS OF THE SPSS ANALYSIS FOR PAGE LAYOUT

Graphics, Images, and Multimedia				
Guidelines	Relative Importance	Paired Samples Statistics Mean (SD) Algerian e-Government & British e-Government	Paired Samples Test Mean (SD)	Significance
1. Use Simple Background Images	4	3.429 (0.202)	-1.285 (0.487)	t=-6.971 p=0.000
		4.714 (0.149)		
2. Label Clickable Images	4	2.286 (0.185)	-2.429 (0.535)	t=-12.021 p=0.000
		4.714 (0.101)		
3. Ensure that Images Do Not Slow Downloads	4	1.856 (0.261)	-3.071 (0.732)	t=-11.103 p=0.000
		4.929 (0.714)		
4. Use Video, Animation, and Audio Meaningfully	4	2.143(0.143)	-1.929 (0.535)	t= -9.546 p=0.000
		4.071 (0.170)		
5. Include Logos	4	4.071 (0.170)	-0.714 (0.394)	t=-4.804 p=0.003
		4.786 (0.101)		
6. Graphics Should Not Look like Banner Ads	4	1.643 (0.143)	-3.357 (0.378)	t=-23.500 p=0.000
		5.000(0.000)		
7. Limit Large Images Above the Fold	4	3.429(0.170)	-1.571 (0.450)	t=-9.242 p=0.000
		5.000 (0.000)		
8. Ensure Web Site Images Convey Intended Messages	4	1.643 (0.210)	-3.214 (0.636)	t=-13.367 p=0.000
		4.857 (0.092)		
9. Limit the Use of Images	3	2.214(0.149)	-2.786 (0.393)	t=-18.735 p=0.000
		5.000 (0.000)		
10. Include Actual Data with Data Graphics	3	3.786(0.184)	-0.929 (0.667)	t=-3.653 p=0.011
		4.714 (0.101)		
11. Display Monitoring Information Graphically	3	1.857 (0.210)	-2.857 (0.475)	t=-15.894 p=0.000
		4.714 (0.149)		

12. Introduce Animation	2	2.571(0.170)	-0.143 (1.029)	t=-3.67 p=0.726
		2.714 (0.240)		
13. Emulate Real-World Objects	2	2.286 (0.185)	-2.071 (0.607)	t=-9.021 p=0.000
		4.357 (0.143)		
14. Use Thumbnail Images to Preview Larger Images	2	1.857(0.210)	-0.786 (0.567)	t=-3.667 p=0.010
		2.642 (0.210)		
15. Use Images to Facilitate Learning	1	2.000 (0.218)	-3.000 (0.577)	t=-13.748 p=0.000
		5.000 (0.000)		
16. Using Photographs of People	1	4.429 (0.130)	2.642 (0.801)	t=8.721 p=0.000
		1.787 (0.1184)		

## V. CONCLUSION

This study sheds light on e-government websites that provide electronic services which facilitate communication between government agencies and citizens. User satisfaction influences web design regardless of the domain, therefore their needs and expectations must be met by the developers. The main subject of this study is the Algerian e-Government which according to the UN has an EGDI Global Rank of 120. This rank was surprising for the researcher considering the population size, infrastructure, and economy of the largest country in Africa. Therefore, this study attempted to identify the reasons that may be rooted to the lack of success of Algerian e-Government services at a global level. For this reason, the UK which ranks 7th in the EGDI Global Rank was selected as the second test subject to serve as a comparative tool to help the researcher identify the underlying reasons for the underperformance of these services. However, this study had to decide on an evaluation tool to use to gather relevant data. Consequently, the researcher identified an ideal set of guidelines known as the RBED&UG which was selected in this study due to its exclusivity and popularity in government agencies as the domain of this topic is e-government usability. The interesting factor regarding these guidelines is that it provides relative importance to each guideline. The researcher was really interested by this feature as the researcher believes that the most impactful guidelines must be addressed before the less important ones. Although these guidelines are used as an evaluation when designing websites, the researcher identified the potential of this to be converted into a questionnaire that could be used by experts when evaluating a system. The complexity of certain aspects of these guidelines, or now proposed questionnaire, suggests that it is not appropriate to be evaluated by normal users, but instead by experts within the field. The researcher believes that this proposal can prove to be very effective as experts are able to rate the chapters and guidelines accordingly, thus the developers can really benefit from their evaluation. There are a total of 18 guidelines in RBED&UG; however, after thorough research and information

gathering, the researcher selected only three categories (Page Layout, Text Appearance, and Graphics, Images & Multimedia) due to their proven role in web usability. This research ultimately aims to explore the reasons that influence the usability of Algerian e-Government services via comparing it with UK e-Government services against a set of guidelines. The results were gathered after an experiment which constituted of an expert evaluation of the two selected systems via completing the proposed questionnaires handed out to them. Their results were inputted into SPSS which provided statistical analysis of the set of data. This data helped identify the mean values of each guideline, their standard deviation, and the significance value of each one. The results display the sheer difference in quality between the Algerian and British e-Government services across all three guideline categories. The three weakest guidelines of each chapter for the Algerian e-Government services include Use Frames when Functions Must Remain Accessible, Highlighting Information, and Graphics Should Not Look like Banner Ads (belonging to Page Layout, Text Appearance, and Graphics, Images & Multimedia respectively). The Algerian e-Government services representatives and developers should focus on these results and improve their systems according to the results of this research, giving most attention to the weakest areas as identified by this study whilst also paying attention to their relative importance. The combination of these two integral factors will help provide effectiveness and efficiency during the evaluation and development process of a system.

## VI. LIMITATION AND FUTURE WORK

This study provides solutions specific to the three usability guideline sections researched. Therefore, there is an opportunity for further research to expand on the remaining 15 guidelines of the RBED&UG. In addition, only two countries were examined in this research.

### REFERENCES

- [1] N. B. Mohammed, and H. B. Abdelhakim. "The success factors of e-government strategy in North Africa: A comparative study between Algerian and Tunisian digital strategy." 2014 4th International Symposium ISKO-Maghreb: Concepts and Tools for knowledge Management (ISKO-Maghreb). IEEE, 2014.
- [2] F. Bélanger, and L. Carter. "Trust and risk in e-government adoption." *The journal of strategic information systems* 17.2 (2008): 165-176.
- [3] M. Benaïda. "Significance of culture toward the usability of web design and its relationship with satisfaction." *Universal Access in the Information Society* 21.3 (2022): 625-638.
- [4] N. Bevan, Nigel, and S. Lonke. "Are guidelines and standards for web usability comprehensive?" *International Conference on Human-Computer Interaction*. Springer, Berlin, Heidelberg, 2007.
- [5] M. Benaïda, Developing Arabic usability guidelines for e-learning websites in higher education. University of Salford (United Kingdom), 2014.
- [6] N. Bevan, "Guidelines and standards for web usability." *Proceedings of HCI International*. Vol. 2005. Lawrence Erlbaum, 2005.
- [7] J. D. Twizeyimana, and A. Annika, "The public value of E-Government—A literature review." *Government information quarterly* 36.2 (2019): 167-178.
- [8] United Nations e-government, E-Government Survey, The Future of Digital Government, 2020.
- [9] S.E. Colesca, and D. Liliana. "Adoption and use of e-government services: The case of Romania." *Journal of applied research and technology* 6 (2008): 204-217.
- [10] G. Hu, W. Pan, M. Lu, and J. Wang, "The widely shared definition of e - Government: An exploratory study." *The Electronic Library* (2009).
- [11] World bank. Accessed 14 August, 2022 Retrieved from <https://www.worldbank.org/en/topic/digitaldevelopment/brief/e-government>.
- [12] ISO 9241-210, International Standard: Ergonomics of Human-System Interaction –Part 210: Human-Centred Design for Interactive Systems, 1st edition 2010-03-15, Reference number ISO 9241-210:2010(E).
- [13] J. V. Dijk, W. Pieterse, A.V. Deuren, and W. Ebbens, "E-services for citizens: The Dutch usage case." *International Conference on Electronic Government*. Springer, Berlin, Heidelberg, 2007.
- [14] Z. Huang, and M. Benyoucef. "Usability and credibility of e-government websites." *Government information quarterly* 31.4 (2014): 584-595.
- [15] M. Benaïda, "Cross-cultural web design and education: a comparison between Arab universities and US universities based on Hofstede cultural dimensions." *IJCSNS* 18.10 (2018).
- [16] P. Verdegem, and G. Verleye, "User-centered E-Government in practice: A comprehensive model for measuring user satisfaction." *Government information quarterly* 26.3 (2009): 487-497.
- [17] V. Venkatesh, M. G. Morris, G.B. Davis, and F.D. Davis, (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.
- [18] J. Kaipio, T. Lääveri, H. Hyppönen, S. Vainiomäki, J. Reponen, A. Kushniruk, E. Borycki, and J. Vänskä, "Usability problems do not heal by themselves: National survey on physicians' experiences with EHRs in Finland." *International Journal of Medical Informatics* 97 (2017): 266-281.
- [19] V. Venkatesh, H. Hoehle, and R. Aljafari. "A usability evaluation of the Obamacare website." *Government information quarterly* 31.4 (2014): 669-680.
- [20] T. Stewart, "Websites—quality and usability." *Behaviour & Information Technology* 31.7 (2012): 645-646.
- [21] N.E. Youngblood, and J. Mackiewicz. "A usability analysis of municipal government website home pages in Alabama." *Government Information Quarterly* 29.4 (2012): 582-588.
- [22] A. F. Al-Bataineh, and S. H. Mustafa. "How Jordanian e-Government websites respond to the needs of people with disabilities." 2016 7th International Conference on Computer Science and Information Technology (CSIT). IEEE, 2016.
- [23] M.Z. Yazid, A.H. Jantan, A.A. Abd Ghani, A. Kamaruddin, and N. Admodisastro, "Accessibility design issues with Malaysian news websites: a case study using a checker and WAVE." *Int J Eng Technol* 7.4 (2018): 69-73
- [24] C. Yingnan, C., Y. Xiaoping, Y., and X. Wang, "The research and realization of the system of the timeliness assessment of E-government web sites based on the semantic analysis." 2011 International Conference on E-Business and E-Government (ICEE). IEEE, 2011.
- [25] Zhao, C., Marghitu, D., & Mou, L. (2016). An Exploratory Study of the Chinese Provincial Government Portals Accessibility. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 1093-1101). Association for the Advancement of Computing in Education (AACE).
- [26] S. F. Verkijika, and L. De Wet. "E-government adoption in sub-Saharan Africa." *Electronic Commerce Research and Applications* 30 (2018): 83-93.
- [27] M. Dollie, and S. Kabanda. "e-Government in Africa: perceived concerns of persons with disabilities (PWDs) in South Africa." *Proceedings of the European Conference on e-Government, ECEG*. 2017.
- [28] S.H. Said, Investigation of usability shortcomings on the user interfaces of web based e-Government systems: a case of Tanzania. Diss. The University of Dodoma, 2015.
- [29] W. Yaokumah, S. Brown, and R. Amponsah. "Accessibility, quality and performance of government portals and ministry web sites: a view using diagnostic tools." 2015 Annual Global Online Conference on Information and Computer Technology (GOICT). IEEE, 2015.

- [30] M. J. Albers, "Tapping as a measure of cognitive load and website usability." Proceedings of the 29th ACM international conference on Design of communication. 2011.
- [31] N.S. Aziz, A. Kamaludin, N.S. Sulaiman, and A. Yacob, "Measuring the Role of Satisfaction in Website Usability Model." *Advanced Science Letters* 24.10 (2018): 7762-7768.
- [32] M. H. Miraz, A. Maaruf, and P. Excell. "Multilingual website usability analysis based on an international user survey." arXiv preprint arXiv:1708.05085 (2017).
- [33] M.Z. Lallmahomed, N. Lallmahomed, and G.M. Lallmahomed, "Factors influencing the adoption of e-Government services in Mauritius." *Telematics and Informatics* 34.4 (2017): 57-72.
- [34] M. Kurfalı, A. Arifoğlu, G. Tokdemir, and Y. Paçin, (2017). Adoption of e-government services in Turkey. *Computers in Human Behavior*, 66, 168-178.
- [35] C. Lodge, "The impact of culture on usability: designing usable products for the international user." *International Conference on Usability and Internationalization*. Springer, Berlin, Heidelberg, 2007.
- [36] A. M. Aladwani, (2013). A cross-cultural comparison of Kuwaiti and British citizens' views of e-government interface quality. *Government Information Quarterly*, 30(1), 74-86.
- [37] Usability testing, Accessed 20 August 2022. Retrieved from <https://www.usability.gov/how-to-and-tools/methods/usability-testing.html>.
- [38] D. Fonseca, and F. J. arcía-Peñalvo. "Interactive and collaborative technological ecosystems for improving academic motivation and engagement." *Universal Access in the Information Society* 18.3 (2019): 423-430.
- [39] H. Huang, and C. W. Chen. "Creating different learning experiences: assessment of usability factors in an interactive three-dimensional holographic projection system for experiential learning." *Universal Access in the Information Society* 18.3 (2019): 443-453.
- [40] M. Fuller-Tyszkiewicz, B. Richardson, B. Klein, H. Skouteris, H. Christensen, D. Austin, ... and A. Ware, "A mobile app-based intervention for depression: end-user and expert usability testing study." *JMIR mental health* 5.3 (2018): e9445.
- [41] J. Krzewińska, A. Indyka-Piasecka, M. Kopel, E. Kukla, Z. Telec, and B. Trawiński, "Usability testing of a responsive web system for a school for disabled children." *Asian Conference on Intelligent Information and Database Systems*. Springer, Cham, 2018.
- [42] F.J. C. De Souza, I.T. Monteiro, and P. M. Jucá, "Game for aNy heuristic evaluation (G4NHE): a generalization of the G4H gamification considering different sets of usability heuristics." *Universal Access in the Information Society* 18.3 (2019): 489-505.
- [43] J. Díaz, J., C. Rusu, C, and C.A. Collazos, "Experimental validation of a set of cultural-oriented usability heuristics: e-Commerce websites evaluation." *Computer Standards & Interfaces* 50 (2017): 160-178.
- [44] P. M. Jucá, I. Teixeira Monteiro, and J. C. D Souza Filho, "Game for heuristic evaluation (G4H): a serious game for collaborative evaluation of systems." *International Conference on Human-Computer Interaction*. Springer, Cham, 2017.
- [45] O. Alhadreti, and P. Mayhew. "To intervene or not to intervene: An investigation of three think-aloud protocols in usability testing." *Journal of Usability Studies* 12.3 (2017): 111-132.
- [46] B. Murillo, S. Vargas, A. Moquillaza, L. Fernández, and F. Paz, "Usability testing as a complement of heuristic evaluation: A case study." *International conference of design, user experience, and usability*. Springer, Cham, 2017.
- [47] A. Atashi, R. Khajouei, A. Azizi, and A. Dadashi,. "User Interface problems of a nationwide inpatient information system: a heuristic evaluation." *Applied clinical informatics* 7.01 (2016): 89-100
- [48] J. Falkowska J. Sobiecki,, and M. Pietrzak. "Eye tracking usability testing enhanced with EEG analysis." *International Conference of Design, User Experience, and Usability*. Springer, Cham, 2016.
- [49] R. Inostroza, C. Rusu, S. Roncagliolo, V. Rusu,, and C.A. Collazos, "Developing SMASH: A set of SMARtphone's uSability Heuristics." *Computer Standards & Interfaces* 43 (2016): 40-52.
- [50] P. Chynał, and J. Sobiecki. "Statistical verification of remote usability testing method." *Proceedings of the Multimedia, Interaction, Design, and Innovation*. 2015. 1-7.
- [51] Bezerra, C., Andrade, R. M., Santos, R. M., Abed, M., de Oliveira, K. M., Monteiro, J. M. & Ezzedine, H. (2014, October). Challenges for usability testing in ubiquitous systems. In *Proceedings of the 26th Conference on l'Interaction Homme-Machine* (pp. 183-188).
- [52] S. M. Ko, W. S. Chang, and Y.G. Ji, "Usability principles for augmented reality applications in a smartphone environment." *International journal of human-computer interaction* 29.8 (2013): 501-515.
- [53] C. Boothe, L. Strawderman, and E. Hosea. "The effects of prototype medium on usability testing." *Applied ergonomics* 44.6 (2013): 1033-1038.
- [54] A. Sivaji, A. Abdullah, and A.G. Downe, "Usability testing methodology: Effectiveness of heuristic evaluation in E-government website development." 2011 fifth Asia modelling symposium. IEEE, 2011.
- [55] T. P. Thyvalikakath, V. Monaco, H. Thambuganipalle and T. Schleyer, "Comparative study of heuristic evaluation and usability testing methods." *Studies in health technology and informatics* 143 (2009): 322.
- [56] A. C. B. Garcia, C. Maciel, and F. B. Pinto, "A quality inspection method to evaluate e-government sites." *International Conference on Electronic Government*. Springer, Berlin, Heidelberg, 2005.
- [57] I. Prazina, E. Kovač, D. Pozderac, Z. Ramić, V. Okanović, and E. Cogo, "Testing Web Page Layouts Using Galen Framework Tests Generated by The Mockup Tool." 2019 27th Telecommunications Forum (TELFOR). IEEE, 2019.
- [58] R. Volentine, A. Owens, C. Tenopir, and M. Frame, "Usability testing to improve research data services." *Qualitative and Quantitative Methods in Libraries* 4.1 (2017): 59-68
- [59] E. Mwangi, S. Kimani, and A. Mindila. "A Review Of Web-Based GIS Usability Elements." *J. Inform. Technol* 4 (2019): 2-13.
- [60] G. W. Tan, and K.K. Wei, "An empirical study of Web browsing behaviour: Towards an effective Website design." *Electronic Commerce Research and Applications* 5.4 (2006): 261-271.
- [61] T. Yang, J. Linder, and D. Bolchini. "DEEP: design-oriented evaluation of perceived usability." *International Journal of Human-Computer Interaction* 28.5 (2012): 308-346.
- [62] S. Djamasbi, M. Siegel, and T. Tullis. "Generation Y, web design, and eye tracking." *International journal of human-computer studies* 68.5 (2010): 307-323.
- [63] A. M. Fiore, and H. J. Jin,. "Influence of image interactivity on approach responses towards an online retailer." *Internet Research* (2003).
- [64] S. Ram, and N. Paliwal. "Design and Development of Multimedia Based User Education Program: The Advantages of YouTube." (2013).

# The Effect of Artificial Neural Network Towards the Number of Particles of Rao-Blackwellized Particle Filter using Laser Distance Sensor

Amirul Jamaludin<sup>1</sup>, Norhidayah Mohamad Yatim<sup>2</sup>, Zarina Mohd Noh<sup>3</sup>

Centre for Telecommunication Research & Innovation (CeTRI)-Fakulti Kejuruteraan Elektronik & Kejuruteraan Komptuer (FKEKK), Universiti Teknikal Melaka (UTeM) Malaysia, Melaka, Malaysia

**Abstract**—Rao-Blackwellized particle filter (RBPF) algorithm aims to solve the Simultaneous Localization and Mapping (SLAM) problem. The performance of RBPF is based on the number of particles. The higher the number of particles, the better the performance of RBPF. However, higher number of particles required high memory and computational cost. Nevertheless, the number of particles can be reduced by using high-end sensor. By using high-end sensor, high performance of RBPF can be achieved and reduced the number of particles. But the development of the robot came at a high cost. A robot can be equipped with low-cost sensor in order to reduce the overall cost of the robot. However, low-cost sensor presented challenges of creating good map accuracy due to the low accuracy of the sensor measurement. For that reason, RBPF is integrated with artificial neural network (ANN) to interpret noisy sensor measurements and achieved better accuracy in SLAM. In this paper, RBPF integrated with ANN is experimented by using Turtlebot3 in real-world experiment. The experiment is evaluated by comparing the resulting maps estimated by RBPF with ANN and RBPF without ANN. The results show that RBPF with ANN has increased the performance of SLAM by 25.17% and achieved 10 out of 10 trials of closed loop map by using only 30 particles compared to RBPF without ANN that needs 400 particles to achieve closed loop map. In conclusion, it shows that, SLAM performance can be improved by integrating RBPF algorithm with ANN and reduces the number of particles.

**Keywords**—SLAM; occupancy grid map; Rao-Blackwellized particle filter; artificial neural network; laser distance sensor

## I. INTRODUCTION

Simultaneous localization and mapping (SLAM) plays an important role in robotics, and particularly in a mobile robot system. SLAM's primary objective is to jointly measure the robot's position as well as map the surrounding of the robot [1]–[5]. The essential of the SLAM algorithm is to map the unknown environment and at the same time exploring the environment. Then, the resulting maps can be used for various applications such as autonomous navigation and search and rescue.

SLAM algorithm commonly used the occupancy grid map (OGM) as a map representation. The maps generate precise metric maps that are close to the detail environmental representations [6]–[8]. This maps require an accurate position of the robot, which makes it a difficult process because of the lack of efficiency of odometric system. The map generated

using raw odometry data only as the position of the robot cannot be sustain and suffers a serious error due to the dead reckoning of odometry system [9]. But these problem can be improved by using Rao-Blackwellized particle filter (RBPF) which can compute more accurate position of the robot. The map is more accurate and globally consistent using RBPF approach. The accuracy of RBPF algorithm is based on the number of particles that required high memory and computational cost [10]–[13]. The higher the number of particles, the more accurate the robot's position and map construction. Therefore it required high memory usage and computational cost to solve the SLAM problem with occupancy grid.

In addition, to achieve high accuracy of occupancy grid map, many notable research have carried out SLAM using high-end sensor to achieve high accuracy map [11], [14]–[18]. Nevertheless, the development of the robot came at a high cost. Therefore, a low-cost sensor could be incorporated to the robot as an alternative to reduce the cost. However, low-cost sensors presented challenges of creating good map accuracy. This is due to the low measurement accuracy of the sensors, which can affect map construction. This problem can be solved by integrating the SLAM technique which is RBPF algorithm with an artificial neural network (ANN) while using low-cost sensors. ANN has been used with occupancy grid maps to interpret noisy sensor measurements and achieved better accuracy of the map [5], [19], [20]. Furthermore, [19] and [5] also shows the number of particles can be reduced with better accuracy of the sensor measurement. Therefore, ANN integrated with RBPF algorithm can reduce the computational cost of the particle filter and achieve better accuracy of OGM. Hence, in this paper, ANN integrated with RBPF algorithm is introduced to improve the sensor measurement and enhance the SLAM technique by using only low-cost sensor.

This paper aims to reduce the number of particles consumption by improving the measurement accuracy of a low-cost laser distance sensor (LDS) and subsequently improve the performance of the SLAM algorithm by incorporating ANN technique with RBPF algorithm. The organization of this paper is as follows: Section 2 reviews the past studies related to the RBPF, sensor measurements accuracy and ANN. Section 3 describes the methodology of the ANN model training and RBPF algorithm framework integrated with ANN. Section 4 analyzes the performance of ANN model after the training and



reports the results of the RBPF algorithm after integrating with ANN. Lastly, section 5 concludes the finding of this paper.

## II. RELATED WORK

The Rao-Blackwellized Particle Filter, also referred to RBPF is employed in the grid-based SLAM algorithm. RBPF is a version of particle filter-based SLAM which is an effective implementation of particle filter-based SLAM. By using Gaussian substructures in the model, RBPF improves the efficiency of the particle filter algorithm [21]. In this approach, RBPF approximates the belief distribution of the robot's pose, while each particle keeps an individual map of the environment, as depicted in Fig. 1. The main contribution of this method is reducing the number of particles used [21]–[24]. This method uses only the most recent observation as well as the most accurate proposal distribution. This can greatly reduce the robot pose uncertainty during filter's prediction step. In addition, the resampling operation is performed selectively in order to address the particle depletion problem. For example in [25], to achieve high accuracy of grid-based SLAM, the classical particle filter required 10000 particles while in [9], RBPF only need 100 particles to achieve high accuracy of SLAM. One of the ROS package implement RBPF algorithm is Gmapping package [26]–[32]. However, since RBPF is based on particle filter-based SLAM, the performance of the algorithm relies on the number of particles. The higher the number of particles, the better is the SLAM performance. This is due to the computation of RBPF itself is based on probability distribution. The higher the number of particles, the higher is the probability to estimate the correct poses and map. As a result, better accuracy of SLAM can be achieved. As in [10], the author shows that with higher number of particles, the map constructed is more accurate than the lower number of particles.

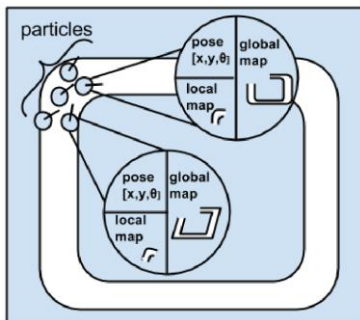


Fig. 1. Each particles contains a hypothesis of robot pose and maintain its own map [10]

A consistent grid-based SLAM algorithm capable to perform loop closure in an unstructured environment. But it requires high precision of the laser scanner to obtain the best hypothesis to build the grid map of the environment. The high precision sensor measurement also can reduce the number of particle usage. For example in [30] and [32], both paper used gmapping package and used the default number of particles which is 30. In [30] the authors successfully achieve high accuracy SLAM by using Hokuyo urg-04lx-ug01 2D lidar which is in high-cost sensor category. While in [32], the map produced is not very accurate since the author only used RGB-D sensor Kinect of Xbox 360 which is low-cost sensor to

perform the SLAM task. But as mention before, low-accuracy SLAM can be mitigated by increasing number of particles. In [10], the author increased the number of particles to 500 to achieve better accuracy SLAM since they used low-cost sensor. However, higher number of particles might suffer from forbidding memory burden and higher computational cost. This problem can be overcome by integrating the SLAM technique with an artificial neural network (ANN) while using low-cost sensors [5], [19], [20], [33]. The noisy dataset from the sensor of the mobile robot are used to train the ANN learner. Afterwards, the ANN model is then applied in the RBPF SLAM algorithm to execute the SLAM task.

In this paper, there are two strategies of dataset that have been reviewed to train the ANN network. Firstly, the training network using the position of each of the grid cells of OGM [5], [19], [33]. Secondly, by using the distance from sensor to obstacles [20]. Firstly, in [19], the author collect the dataset from the front six infrared sensor of E-puck mobile robot as depicted in Fig. 2. Then, the dataset is used as the input layer of the ANN training as shown in ANN configuration in Fig. 3. This ANN is trained to estimate the occupancy probability value  $x, y$  position of the cell. The same architecture of ANN is used in [5], [33]. The authors used the front four sensors' measurements of Khepera III robot that are closest to the cell position as the input of ANN. The same target OGM output of the ANN model which is  $x$  and  $y$  position of the OGM's cell ( $C_{xy}$ ) are used to train the input as shown in Fig. 3.

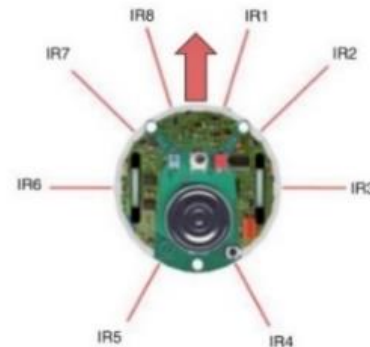


Fig. 2. E-puck mobile robot is equipped with eight infrared sensors [19]

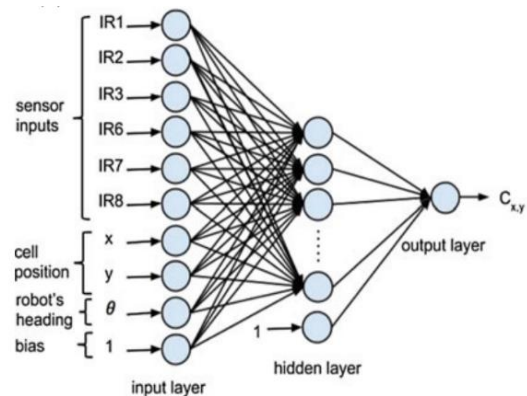


Fig. 3. A nine-input ANN consisting of infrared sensors, cell position and heading of the robot [19]

While in [20], the dataset is collected based on the reading of the sensor values that contain the estimated distance to obstacle. Then this dataset is trained using ANN to correct the distance measurement to the obstacles by using the real distance as the target output of the ANN. The grid map for the observed region is then built using this corrected measurement. To compute the grid map estimation in this region, only two values are taken into account: 0 for open space and 1 for occupied space (i.e., objects in the space). In the context of computation time, estimation of ANN by using distance to the obstacles is considered faster [20] than estimation by using each of the grid cells [5], [19], [33]. This is due to the estimation by distance does not evaluate cell by cell and only taking value 0 and 1 as mention above. While in the latter approach, the estimation of the map computes each cell of the region by calculating  $x$  and  $y$  cell's occupancy value of the OGM. Cell by cell evaluation can cause slow computation time especially in large environment that required many cells to build the OGM. Hence, real time implementation is not feasible. Therefore, ANN based on the distance is implemented in this paper since it is faster and more suitable for real-time implementation due to the computation does not require for each grid cell but for the entire detected region. The presented methods of ANN training are summarized in Table I.

TABLE I. COMPARISON OF ANN TRAINING METHOD

ANN training method	Input/Output	Pros/Cons
Based on cell-by-cell OGM	Input: Multiple sensor measurement value, cell position $(x, y)$ and robot heading, $\theta$ Target output: Occupancy probability value	Pros: Exploit adjacent sensor measurements and less depending on accuracy of the sensor Cons: slow computation
Based on distance	Input: Sensor measurement value Target output: Actual distance	Pros: Fast computation Cons: more depending on accuracy of the sensor

### III. METHODOLOGY

For the methodology phase, the operation is divided into four phases which is sensor data collection, ANN training, ANN model integrated with RBPF algorithm and lastly, evaluation of the resulting map. Firstly, the LDS data points from the laser distance sensor LDS-01 are collected. The data points are collected from the real-world sensor. These data are collected between the range measurement of 0.12 to 3.5m as shown in Fig. 4(a). At each interval of 0.1m, 2000 measurements were collected. Since LDS-01 sensor is capable to sense 360 degree of the surrounding, the data are taken only at 0 degree. This is because only the measurement at 0 degrees is perpendicular to the wall as shown in Fig. 4(b). The measurements at other degrees facing the wall would not be precisely equal to the ground truth but at longer range as they are slightly slanted towards the wall. Then, these data were used as the input of the ANN training. The actual distance between the mobile robot and the wall was used as the reference or target output.

After the data collection, these data are trained to build ANN model in the second phase. This ANN training employs a multilayer feed-forward network. The architecture of the

network is made up of an input layer, a hidden layer and a single neuron for output layer as depicted in Fig. 5. In this paper, a tangent sigmoidal activation function and a linear transfer function were used for the hidden and output layers, respectively.

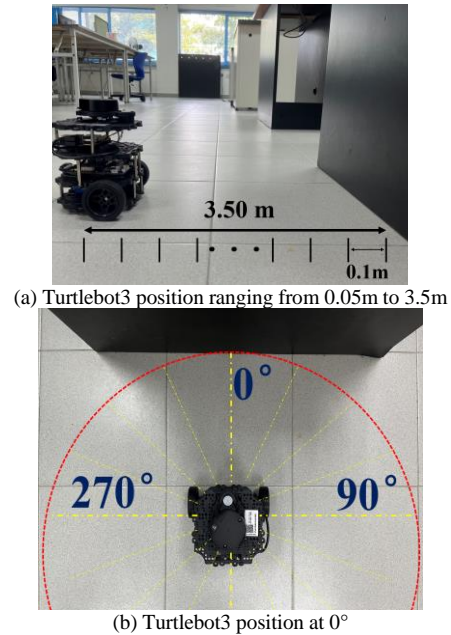


Fig. 4. Position of the turtlebot3 to the wall in real-world experiment

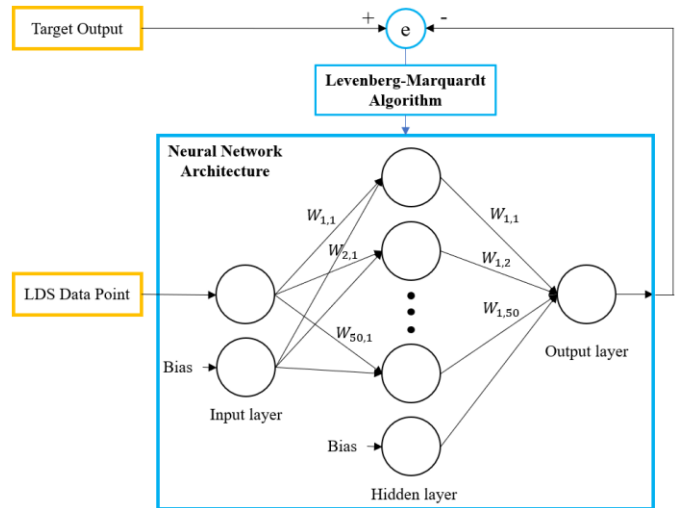


Fig. 5. Architecture of ANN model training

After the ANN structure is established, the next step is training the model. A set of input and reference data pairs as well as a training rule are presented during the training procedure. As mention in the first phase, the input is the LDS data point that have been collected and the reference data is the actual distance of the sensor to the wall. The ANN generates its own output from the input and compares it to the reference data. To ensure the difference of the output is small as possible to the reference value, the interconnection weights,  $W$  between the nodes are determined by using training rule. The challenge with learning is finding the optimal weights  $W$  combination

and attempts to minimize the mean-square-error (MSE) between the reference value and the predicted output. The use of a backpropagation algorithm is the most widely used training rule for error minimization. An improved backpropagation algorithm, the Levenberg Marquardt (LM) algorithm, come out to be a faster and more efficient approach for training medium-sized feedforward neural networks [34]–[37]. Therefore, in this paper, the LM algorithm is employed. The output of the network is also computed based on the number of neurons in the hidden layer. The number of neurons in the hidden layer is determined by the response of the output that resulted in the smallest MSE error. At the end of the training phase, 50 number of neurons is selected as the number of neurons which has managed to achieve MSE value of  $9.55 \times 10^{-5}$ .

After the desired ANN model is obtained, the next phase is to integrate the model with Rao-Blackwellized particle filter (RBPF) algorithm as shown in Fig. 6. The RBPF algorithm implemented in Gmapping package of Robot Operating System (ROS) is used in this paper. RBPF requires odometry data as well as sensor observations to solve grid-based SLAM problem. The new sensor observation that has been improved using ANN is integrated in this phase. The main idea of RBPF is to use odometry data,  $u_{1:t-1}$  and observations,  $z_{1:t}$  to estimate the trajectories of robot,  $x_{1:t}$  and map,  $m$ . This joint posterior, which is written as  $p(x_{1:t}, m|z_{1:t}, u_{1:t-1})$  can be factored as follows:

$$p(x_{1:t}, m|z_{1:t}, u_{1:t-1}) = p(m|z_{1:t}, x_{1:t}) \cdot p(x_{1:t}|z_{1:t}, u_{1:t-1}) \quad (1)$$

The computations are made simpler by factorization, which enables the process to be executed in two steps. First, odometry and observation data can be used to estimate the robot's trajectory,  $p(x_{1:t}|z_{1:t}, u_{1:t-1})$ . Once the  $x_{1:t}$  and  $z_{1:t}$  are known, the map  $p(m|z_{1:t}, x_{1:t})$  can be computed.

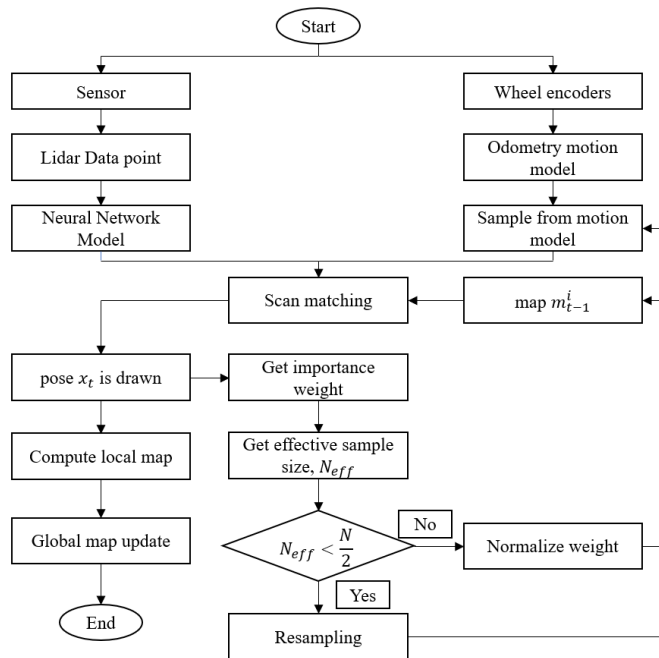


Fig. 6. RBPF algorithm integrated with ANN flow chart

The estimation is more complex with respect to  $p(x_{1:t}|z_{1:t}, u_{1:t-1})$ . A particle filter is used for this purpose where any particle reflects the robot's potential trajectory. Based on the potential trajectory, an individual map for each particles can be computed. Then, the full map is built by the corresponding particles. The sampling importance filter (SIR) is commonly used in this algorithm. The SIR function is used to select the particles with the highest probability and the output of the algorithm is the the associated map. In order to improve the algorithm, the Rao-Blackwellized SIR filter uses the most recent sensor observations and odometry readings where available. Specifically, in RBPF, the approximation of the trajectory  $p(x_{1:t}|z_{1:t}, u_{1:t-1})$  was remodeled to  $p(x_{1:t}|m_{t-1}^i, x_{t-1}^i, z_{1:t}, u_{1:t-1})$  [31]. Scan matching is used to match the observations to the map that has been created so far, optimize observation probabilities, and provide information about the most likely poses of the robot. In this step, accuracy of the sensor measurement is important to get better observation for scan matching step and obtain better pose estimation.

After that, particles are resampled according to their weight. Particles with higher weight will be most likely to be resampled for the next generation. All particles have the same weight after resampling. To avoid the resampling from removing good particles, a careful resampling step is taken. Hence, a selective resampling technique which is effective samples size,  $N_{eff}$  is proposed which is to decide when to perform a resampling step and is described as:

$$N_{eff} = \frac{1}{\sum_{i=1}^N (\omega^i)^2} \quad (2)$$

where  $\omega^i$  represents the normalized weight of  $i$ -th particle. The weights of the samples are approximately equal if they are close to the target distribution. As the samples deviate from the target distribution, their weights variance increases and  $N_{eff}$  decreases. Every time  $N_{eff}$  falls below  $N/2$ , the resampling procedure is initiated (where  $N$  is the number of particles used in the filter). This greatly reduces the risk of replacing useful particles, as resampling is performed only when necessary and the cumulative number of such operations is reduced.

Last phase is to evaluate the performance of RBPF integrated with ANN in real-world experiment. The RBPF with ANN will be compared to RBPF without ANN. The maps that have been obtained from both algorithms respectively will be evaluated with ground truth map using number of inliers evaluation. To evaluate the resulting maps quantitatively, similarities between both maps are measured using number of inliers that are obtained from RANSAC algorithm. RANSAC algorithm can calculate the similarity point (inliers) between resulting maps and ground truth map. The higher the number of inliers, the better the similarity of the resulting map and ground truth map and the better the performance of the map constructed. Ground truth map is obtained by using RBPF algorithm with high number of particles. This is due to the high number of particles can obtained high accuracy of map. In this paper, 1000 particles is used to obtain the ground truth map. Additionally, the robot explored the environment in smaller loops. This process is repeated until a satisfactory ground truth map is obtained. By exploring in smaller loops, the

accumulated error of robot's state estimate is kept to the minimum and robot's trajectory maintain on the correct path and did not diverged. Turtlebot3 robot platform is used in this paper to navigate the environment as shown in Fig. 7. The experiment was conducted at Faculty of Electronics and Computer Engineering Universiti Teknikal Malaysia Melaka. The robot was set to explore wing A of the faculty within the red rectangle by following the path marked using red arrow. The size of this real-world environment is approximately 43 x 16 meters. The Turtlebot3 explored the environment using the teleop operation. The average speed of the robot was about 0.12 m/s and simultaneously recorded by the Rosbag tool. Rosbag tool allows recording and playing back the data of the robot that have been recorded. Hence, the recorded data which is the same data can be applied to other algorithms. This way, RBPF algorithm integrated with ANN and without ANN can be evaluated in equal condition. After that, the RBPF with ANN algorithm was observed if there is any improvement after the ANN integration.



Fig. 7. Layout of wing A of faculty of electronics and computer engineering

#### IV. RESULT

The result is divided into two sections. In the first section, ANN model that have been trained is analyzed by comparing the sensor measurement with ANN and without ANN. Afterwards, in the second section, the RBPF algorithm after integrating with ANN is analyzed by comparing the resulting maps with the ground truth map.

##### A. Performance of ANN

The training of the ANN model has managed to achieve MSE value of  $9.55 \times 10^{-5}$ . 50 number of neurons are selected as it has achieved the lowest MSE value. The synaptic weights in each layer are computed and adjusted according to the lowest mean-squared error (MSE). The training automatically stops when generalization stops improving, as indicated by the lowest MSE of the training samples. Finally, weight and bias vectors for this ANN training were determined after several attempts and ANN model is established.

Afterwards, the ANN model that have been established is analyzed. The performance of the model is analyzed based on the comparison of the sensor measurement before (without ANN) and after the ANN model (with ANN) is used. The data of both sensor measurement for every interval of 0.1m were averaged and plotted in the graph respectively as shown in Fig. 8. The graph shows that the sensor reading without ANN

(green line) is slightly different from the actual distance (red line). The difference of the sensor measurement without ANN and actual distance can be observed in more detail in Fig. 9. The graph shows that the error of sensor measurement without ANN to the actual distance is consistently higher. It is also observed that when the range of the distance increases, the distribution of sensors measurements is wider. Fig. 10 shows the histogram of the LDS sensor measurement at 3 meter distance is wider than the histogram of sensor measurement at 1 meter distance as depicted in Fig. 11. Therefore, the standard deviation of the sensor measurements also increases as the distance increase as shown in Fig. 13.

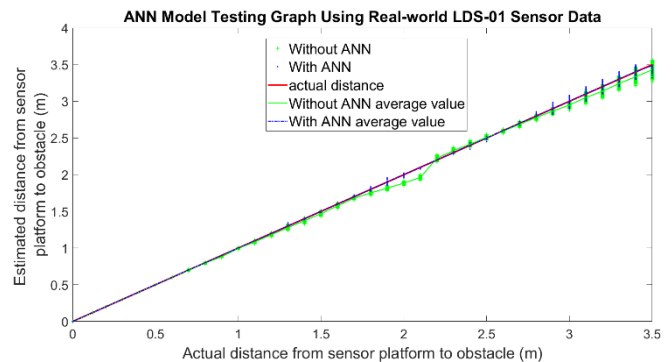


Fig. 8. Graph illustration for the data of ANN model testing using real-world LDS-01 sensor measurements data

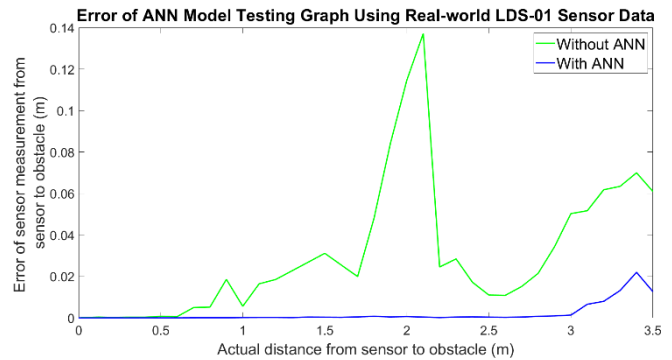


Fig. 9. Comparison of error of sensor measurement without ANN (green) and with ANN (blue)

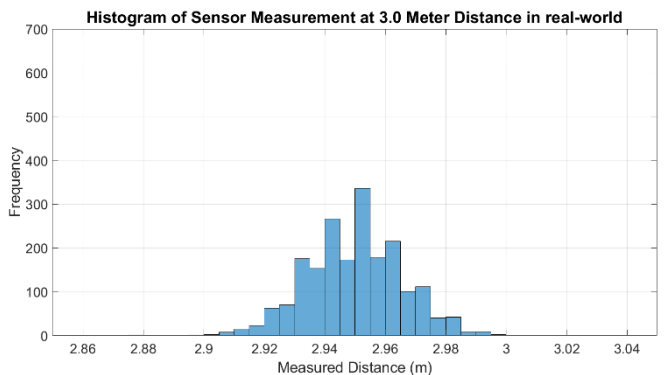


Fig. 10. Histogram of LDS sensor measurement at 3 meter distance

After the LDS sensor data has been applied to the ANN model that have been trained, the value of the data with ANN

(blue line) and the actual distance value (red line) has minimal difference such that the blue and red line are in the same axis as shown in Fig. 8. The error of the sensor measurement also decreased remarkably as shown in Fig. 9. It shows that the non-linearity and the error of the sensor readings are significantly reduced by using ANN. In addition, the histogram of sensor measurement at 3 meter distance using ANN is narrower than the histogram without ANN as shown in Fig. 12. The standard deviation of the sensor measurement with ANN (blue line) is also consistently lower than the standard deviation of the sensor measurement without ANN (green line) as shown in Fig. 13. From these results, it shows that there is a significant improvement of the accuracy of the LDS sensor measurements.

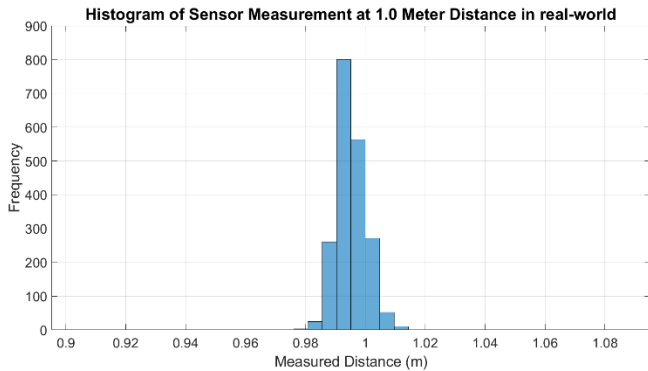


Fig. 11. Histogram of LDS sensor measurement at 1 meter distance (without ANN)

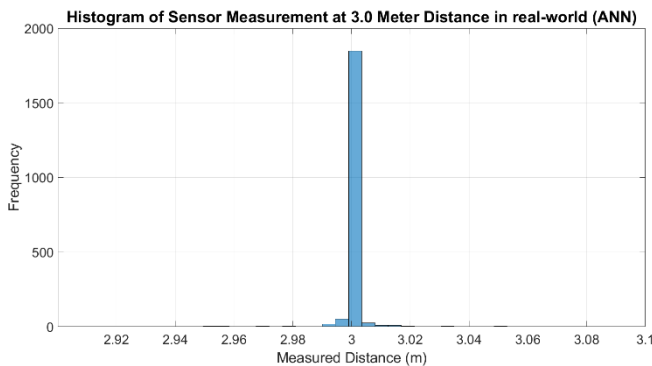


Fig. 12. Histogram of LDS sensor measurement at 3 meter distance (with ANN)

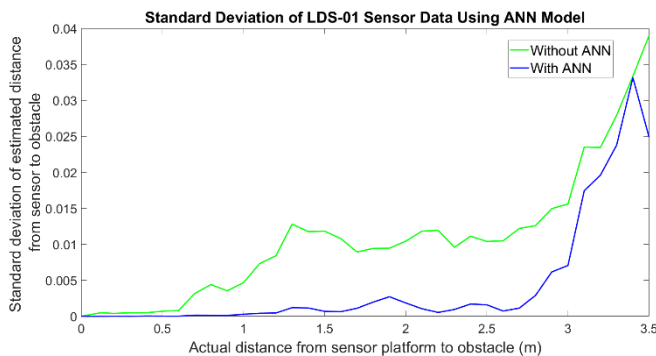


Fig. 13. Standard deviation of LDS-01 sensor data using ANN model before and after testing

### B. Performance of RBPF Algorithm

In this experiment, the number of particles of Gmapping package which is 30 to 400 particles is used. The maps are obtained in the Occupancy Grid map (OGM) representation and saved in pgm format using map saver from map server package. The size of each grid cells of the map is set to  $5\text{ cm}^2$ . In OGM, the black cells are considered occupied, white cells are considered free cells and grey cells are the unknown region that has not been explored yet as shown in Fig. 14. The data that have been recorded is tested 10 times resulting 10 grid maps. Then, the maps that have been obtained using ANN and without ANN are evaluated using number of inliers evaluation to ensure for a consistent result. The dimension of the map is  $1056 \times 608$  cells. Since, the size of the grid cell is  $5\text{ cm}^2$ , this make the size of the real environment in Fig. 14 is approximately  $52.8\text{ m} \times 30.4\text{ m}$ . Within this map, the explored area travelled by Turtlebot3 mobile robot is approximately  $43\text{ m} \times 16\text{ m}$  as stated in Section III.

Primarily, the particles used for the first experiment is 30. From the resulting maps, it is observed that, RBPF algorithm that is not integrated with ANN cannot achieved closed loop condition out of 10 trials as shown in Fig. 15. Most of the resulting maps of RBPF algorithm without ANN show that the algorithm cannot recognize the revisited environment at the loop closure point (green circle) as shown in Fig. 15. This is due to the low accuracy of the LDS sensor measurement reading and consequently effects the RBPF algorithm. To achieve all closed loop condition without using ANN, further experiment is carried out by increasing number of particles. After several trials with increasing number of particles (100, 200 and 400), 400 particles used in the RBPF algorithm without ANN managed to attain loop condition for all trials which is 10 out of 10 as shown in figure 16. The detail results of RBPF with number of particles used are included in Table II. As shown in the table, for 100 and 200 particles, the resulting maps have achieved 4 and 6 closed loop condition out of 10 trials, respectively.

TABLE II. RESULTING MAP CONDITION ACCORDING TO NUMBER OF PARTICLES

SLAM Algorithm	Number of particles	Closed loop map out of 10 trials	Average number of inliers
RBPF without ANN	30	0/10	53
	100	4/10	47.9
	200	6/10	48.5
RBPF with ANN	400	10/10	46.9
	30	10/10	58
	100	10/10	58.3
	200	10/10	61.6
	400	10/10	67.8

It shows that by increasing the number of particles, the accuracy of the grid map can be improved. However, higher number of particles may suffer from limitation of memory capacity and higher computational cost.

Fig. 17 and 18 represent the computational consumption diagrams obtained by the system monitor of Ubuntu for 30 and 400 particles of RBPF algorithm, respectively. The specifications of the computer used are AMD Ryzen 5 3600 GPU 6-core processor with 16GB of RAM. The data is taken after one minute of the execution of one of the tests. It can be observed that RBPF with 400 particles consumed up to 100% of the most cores available during the lapse while 30 particles only consume around 30 to 40% of the cores available. This shows that, higher computational cost is needed to execute higher number of particles.

However, ANN can overcome this problem as mention before by improving the LDS sensor measurement. The ANN model integrated with RBPF algorithm (using 30 particles) is tested 10 times. From the results, it is observed that, RBPF algorithm integrated with ANN have achieved all closed loop map condition which is 10 times out of 10 trials by using only 30 particles. Sensor reading by using ANN has increased the measurement likelihood hence resulting a more accurate estimate in robot localization in the RBPF algorithm. As the localization error has decreased, the robot is able to recognize the revisited area. RBPF with ANN also has proven to achieve this condition with less number of particles more consistently. As a result, we can conclude that the particle consumption can be reduced by integrating ANN with the RBPF algorithm and overcome the computational cost suffers from the high number of particles remarkably. Further experiment by increasing number of particles (100, 200 and 400) is also tested by using ANN. The performance of resulting maps is increased and all resulting maps achieved closed loop condition. As mention in section III, the performance of the resulting maps is compared with the ground truth map (Fig. 14) by using number of inliers. The number of inliers is stated in Table I along with the number of particles.

Then, Fig. 19 plots the performance of the resulting maps by using the number of inliers (y-axis) vs the number of particles (x-axis). As shown in the graph, by using ANN integrated with RBPF, the performance of the resulting maps (red) is consistently higher than the resulting maps without ANN (blue). The performance using ANN is increasing with the number of particles while for RBPF without ANN, the performance is decreasing. We believe that the performance without ANN is decreasing because of the dimension of the resulting maps (without ANN) is less similar to the ground truth map. By using 400 particles (without ANN) better pose estimation is achieved compared to lesser particles. Due to that, closed-loop map condition is achieved when the robot explored the environment and encountered the initial position. Although closed loop map condition is achieved, the dimension of the resulting map is not consistent due to the nonlinearity error of the LDS sensor measurement. As shown in the graph, 30 and 200 particles have achieved higher number of inliers, while 100 and 400 particles achieved lower number of inliers. Due to that, the performance of the resulting maps is not consistent

along the number of particles. Fig. 20 shows one of the examples of the number of inliers point between resulting map (without ANN) and ground truth map by using 400 particles. The dimension of the resulting map (without ANN) is 15.7 m x 42.6 m while ground truth map is 15.9 m x 43 m. Since the dimension of the map is slightly smaller to the ground truth map, hence, only 42 number of inliers point is obtained.

### C. Discussion

It is observed that when using ANN, the quality of the resulting maps is increasing with the number of particles as expected. This is due to the more accurate sensor measurements. Better sensor measurement improved the observation of the robot for scan matching step and obtained better pose estimation as mentioned in Section III. Due to that lower number particles can be used to achieve higher performance of the map building and closed loop condition. For this experiment, only 30 particles are used to accomplished the closed loop condition. Besides, better resulting map is achieved and more similar to the ground truth map by increasing the number of particles using ANN compared to without ANN. As shown in Fig. 21, the number of inliers point between the resulting maps (with ANN) and ground truth map is higher which is 77 inliers compared to without ANN which is 42 as shown in Fig. 20. This is due to the dimension of the map (with ANN) which is 15.95 m x 43.10 m is more similar to the ground truth map (15.9 m x 43 m) compared to without ANN (15.7 m x 42.6 m). Based on the average number of inliers, the overall performance of the resulting maps by using RBPF with ANN has increased by 25.17%. In conclusion, with better accuracy of the sensor measurement integrated with ANN reduces the number of particles consumption. Therefore, ANN can reduce the computational cost of the RBPF algorithm and achieve better accuracy of OGM by using only low-cost sensor.

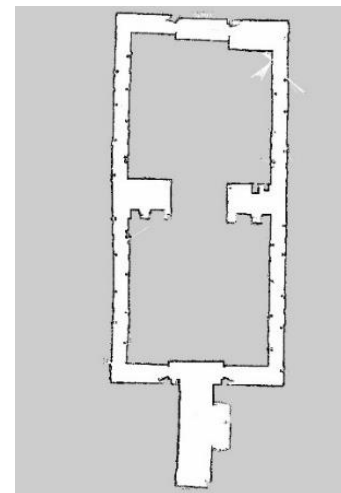


Fig. 14. Ground truth map of the real-world environment



Fig. 15. Non-closed loop map without ANN

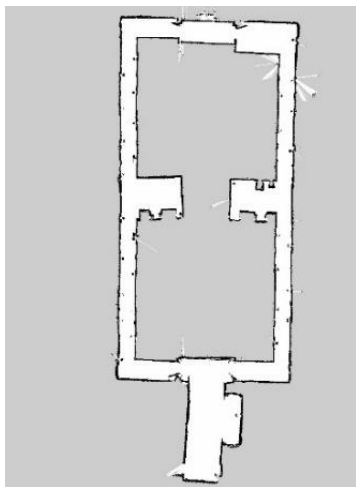


Fig. 16. Closed loop map with ANN

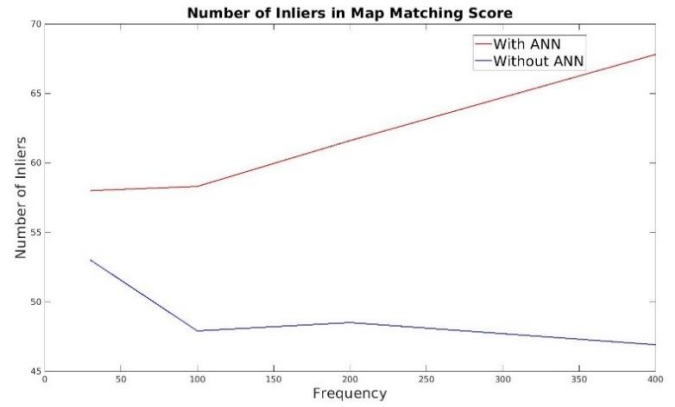


Fig. 19. Map performance by using number of inliers between resulting map and ground truth map

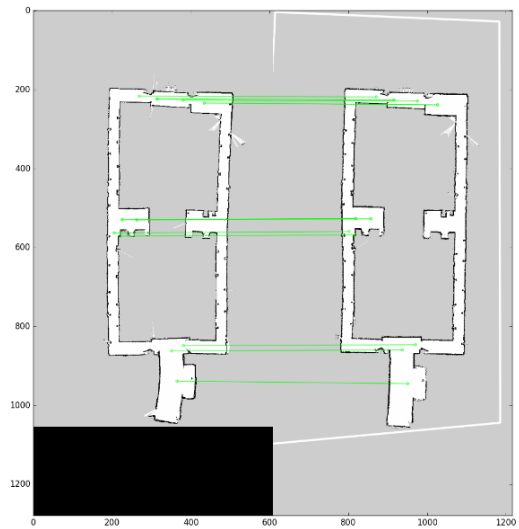


Fig. 20. 42 number of inliers points between resulting map (without ANN) and ground truth map

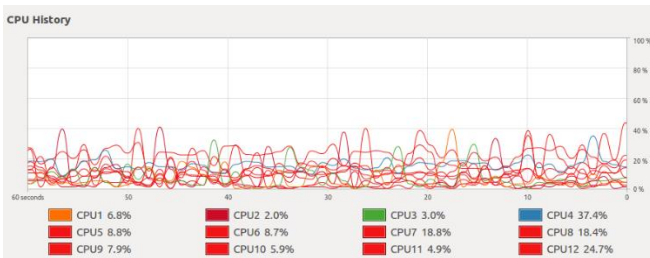


Fig. 17. CPU consumption in one minute of the RBPf using 30 particles.

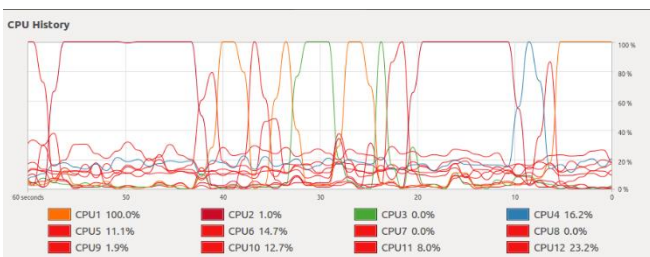


Fig. 18. CPU consumption in one minute of the RBPf using 800 particles

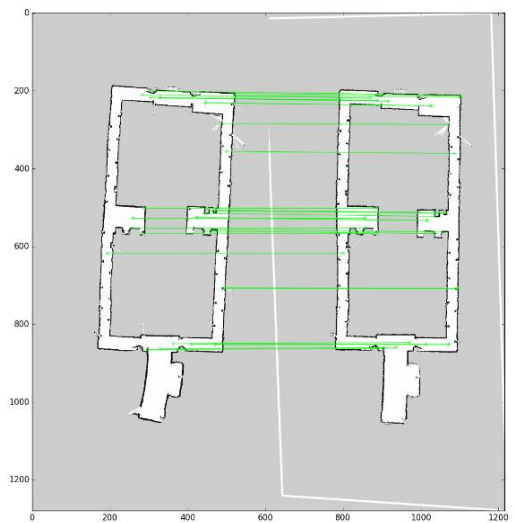


Fig. 21. 77 number of inliers points between resulting map (with ANN) and ground truth map

## V. CONCLUSION

This paper presented the effects of the ANN towards the number of particles of RBPF algorithm using laser distance sensor LDS-01. An investigation has been made to observe the performance of the RBPF algorithm by using ANN and without ANN. The performance of the algorithm is calculated based on the accuracy of the map constructed by both algorithms. Based on the results, RBPF with ANN has achieved all closed loop condition of the resulting maps which is 10 times out of 10 trials by using only 30 particles compared to RBPF without ANN that needs 400 particles to achieved all closed loop condition. Furthermore, the number of inliers of the resulting maps by using RBPF with ANN is higher than RBPF without ANN consistently for the number of particles ranging from 30 to 400 particles and increase the performance by 25.17%. From the results, it can be concluded that using ANN improves the performance of the RBPF algorithm and reduces the number of particles consumption for mobile robot platform with low-cost sensors. For the future works is to test the SLAM algorithm integrated with ANN on cooperative robot. Cooperative robot SLAM has many advantages, including the ability to complete missions quicker and being resilient to the malfunction of any one of the robots. To obtain high accuracy of the cooperative SLAM, many researchers chose to adopt high-end sensor such as 2D SICK and Hokuyo Lidar as robot's perception. But this will cause staggering cost for cooperative robot system. Thus, the alternative is to implement low-cost sensors with limited sensing to perform cooperative SLAM.

## ACKNOWLEDGMENT

We (the authors) would like to acknowledge and express our gratitude to Centre for Research and Innovation Management (CRIM), Centre for Telecommunication Research & Innovation (CeTRI), Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer (FKEKK), Universiti Teknikal Malaysia Melaka (UTeM) and the Ministry of Higher Education Malaysia for the grant RACER/1/2019/TK04/UTEM//5 for providing us with the opportunity, facilities, and finances to conduct this study. We also would like to thank Mr. Hee Thien Tack, Mrs. Fatini Kamaruzaman and Mr. Mahadi Salleh for the technical support during this research.

## REFERENCES

- [1] S. Wen, J. Chen, X. Lv, and Y. Tong, "Cooperative simultaneous localization and mapping algorithm based on distributed particle filter," *Int. J. Adv. Robot. Syst.*, vol. 16, no. 1, pp. 1–8, 2019.
- [2] I. Ullah, X. Su, X. Zhang, and D. Choi, "Simultaneous Localization and Mapping Based on Kalman Filter and Extended Kalman Filter," *Wirel. Commun. Mob. Comput.*, vol. 2020, 2020.
- [3] S. Saeedi, M. Trentini, H. Li, and M. Seto, "Multiple-robot Simultaneous Localization and Mapping - A Review 1 Introduction 2 Simultaneous Localization and Mapping: problem statement," *J. F. Robot.*, vol. 33, no. 1, pp. 3–46, 2016.
- [4] Y. Zhao, T. Wang, W. Qin, and X. Zhang, "Improved Rao-Blackwellised particle filter based on randomly weighted particle swarm optimization," *Comput. Electr. Eng.*, vol. 71, no. August, pp. 477–484, 2018.
- [5] N. M. Yatim, A. Jamaludin, and Z. M. Noh, "Comparison of Sampling Methods for Rao-blackwellized Particle Filter with Neural Network," *Symp. Intell. Manuf. Mechatronics. Symp. 2021. Lect. Notes Mech. Eng. B. Ser. (LNME).*, pp. 60–75, 2021.
- [6] S. M. J. Alexandre, "MRSLAM – Multi-Robot Simultaneous Localization and Mapping," Master's Thesis, no. September, pp. 1–67, 2013.
- [7] M. P. De Melo, L. Cambuim, and E. Barros, "Occupancy Grid Map Estimation Based on Visual SLAM and Ground Segmentation," 2021 Lat. Am. Robot. Symp. 2021 Brazilian Symp. Robot. 2021 Work. Robot. Educ. LARS-SBR-WRE 2021, pp. 288–293, 2021.
- [8] R. Singh and K. S. Nagla, "Improved 2D laser grid mapping by solving mirror reflection uncertainty in SLAM," *Int. J. Intell. Unmanned Syst.*, vol. 6, no. 2, pp. 93–114, 2018.
- [9] D. Hähnel, W. Burgard, D. Fox, and S. Thrun, "An Efficient FastSLAM Algorithm for Generating Maps of Large-Scale Cyclic Environments from Raw Laser Range Measurements," *IEEE Int. Conf. Intell. Robot. Syst.*, vol. 1, no. November, pp. 206–211, 2003.
- [10] N. M. Yatim and N. Buniyamin, "Particle filter in simultaneous localization and mapping (Slam) using differential drive mobile robot," *J. Teknol.*, vol. 77, no. 20, pp. 91–97, 2015.
- [11] H. Jo, H. M. Cho, S. Jo, and E. Kim, "Efficient Grid-Based Rao-Blackwellized Particle Filter SLAM with Interparticle Map Sharing," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 2, pp. 714–724, 2018.
- [12] T. Zhang, C. Wang, Z. Yuan, and M. Zheng, "Improved grid mapping technology based on Rao-Blackwellized particle filters and the gradient descent algorithm," *Syst. Sci. Control Eng.*, vol. 7, no. 1, pp. 65–74, 2019.
- [13] H. Zhou, Z. Yao, C. Fan, S. Wang, and M. Lu, "Rao-Blackwellised particle filtering for low-cost encoder/INS/GNSS integrated vehicle navigation with wheel slipping," *IET Radar, Sonar Navig.*, vol. 13, no. 11, pp. 1890–1898, 2019.
- [14] U. Lidars, M. Torres-torriti, P. Nazate-burgos, F. Paredes-lizama, and J. Guevara, "Passive Landmark Geometry Optimization and Evaluation for Reliable Autonomous Navigation in Mining Tunnels Using 2D Lidars," *Sensors*, vol. 22, no. 8, 2022.
- [15] D. Wu, Y. Meng, K. Zhan, and F. Ma, "A LIDAR SLAM Based on Point-Line Features for Underground Mining Vehicle," *Proc. 2018 Chinese Autom. Congr. CAC 2018*, pp. 2879–2883, 2019.
- [16] G. S. Martins, D. Portugal, and R. P. Rocha, "mrgs: A Multi-Robot SLAM Framework for ROS with Efficient Information Sharing," *Robot Oper. Syst. (ROS), Stud. Comput. Intell.*, vol. 895, pp. 45–75, 2021.
- [17] F. Demim, A. Nemra, K. Louadj, M. Hamerlain, and A. Bazoula, "Cooperative SLAM for multiple UGVs navigation using VSVF filter," *Automatika*, vol. 58, no. 1, pp. 119–129, 2017.
- [18] S. Saeedi, L. Paull, M. Trentini, M. Seto, and H. Li, "Map merging for multiple robots using Hough peak matching," *Rob. Auton. Syst.*, vol. 62, no. 10, pp. 1408–1424, 2014.
- [19] N. M. Yatim and N. Buniyamin, "Development of Rao-Blackwellized Particle Filter ( RBPF ) SLAM Algorithm Using Low Proximity Infrared Sensors," 9th Int. Conf. Robot. Vision, Signal Process. Power Appl., no. September 2017, pp. 395–405, 2017.
- [20] G.-G. Kim, Heon-hui Ha, Yun-Su Jin, "A Study on the Environmental Map Building for a Mobile Robot Using Infrared Range-finder Sensors," *Intell. Robot. Syst. 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ Int. Conf.*, vol. 1, no. October 2003, pp. 711–716, 2003.
- [21] Y. Dai and M. Zhao, "Grey Wolf Resampling-Based Rao-Blackwellized Particle Filter for Mobile Robot Simultaneous Localization and Mapping," *J. Robot.*, vol. 2021, 2021.
- [22] G. Vallicrosa and P. Ridao, "H-SLAM: Rao-Blackwellized particle filter SLAM using Hilbert Maps," *Sensors (Switzerland)*, vol. 18, no. 5, 2018.
- [23] A. S. Aguiar, F. N. dos Santos, H. Sobreira, J. B. Cunha, and A. J. Sousa, "Particle filter refinement based on clustering procedures for high-dimensional localization and mapping systems," *Rob. Auton. Syst.*, vol. 137, p. 103725, 2021.
- [24] N. Akai, L. Y. Morales, and H. Murase, "Simultaneous pose and reliability estimation using convolutional neural network and Rao-Blackwellized particle filter," *Adv. Robot.*, vol. 32, no. 17, pp. 930–944, 2018.



- [25] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, "Robust Monte Carlo localization for mobile robots," *Artif. Intell.*, vol. 128, no. 1–2, pp. 99–141, 2001.
- [26] C. S. Gurel, "Real-Time 2D and 3D Slam Using Rtab-Map , Gmapping , and Cartographer Packages," no. August, 2018.
- [27] M. Islam, T. Ahmed, A. T. Bin Nuruddin, M. Islam, and S. Siddique, "Autonomous Intruder Detection Using a ROS-Based Multi-Robot System Equipped with 2D-LiDAR Sensors," 2020 IEEE Int. Symp. Safety, Secur. Rescue Robot. SSR 2020, pp. 326–333, 2020.
- [28] B. Hampton, A. Al-hourani, B. Ristic, and B. Moran, "RFS-SLAM Robot: An Experimental Platform for RFS Based Occupancy-Grid SLAM," 2017 20th Int. Conf. Inf. Fusion.
- [29] Y. Alborzi, B. S. Jalal, and E. Najafi, "ROS-based SLAM and Navigation for a Gazebo-Simulated Autonomous Quadrotor," 2020 21st Int. Conf. Res. Educ. Mechatronics, REM 2020, 2020.
- [30] R. Yagfarov, M. Ivanou, and I. Afanasyev, "Map Comparison of Lidar-based 2D SLAM Algorithms Using Precise Ground Truth," 2018 15th Int. Conf. Control. Autom. Robot. Vision, ICARCV 2018, pp. 1979–1983, 2018.
- [31] K. Kamarudin, S. M. Mamduh, A. Y. Md Shakaff, and A. Zakaria, "Performance analysis of the microsoft kinect sensor for 2D simultaneous localization and mapping (SLAM) techniques," *Sensors (Switzerland)*, vol. 14, no. 12, pp. 23365–23387, 2014.
- [32] M. Rojas-Fernandez, D. Mujica-Vargas, M. Matuz-Cruz, and D. Lopez-Borreguero, "Performance comparison of 2D SLAM techniques available in ROS using a differential drive robot," 2018 28th Int. Conf. Electron. Commun. Comput. CONIELECOMP 2018, vol. 2018-Janua, pp. 50–58, 2018.
- [33] N. A. Yatim, N. Buniyamin, Z. M. Noh, and N. A. Othman, "Occupancy grid map algorithm with neural network using array of infrared sensors," *J. Phys. Conf. Ser.*, vol. 1502, no. 1, 2020.
- [34] S. Mammadli, "Financial time series prediction using artificial neural network based on Levenberg-Marquardt algorithm," *Procedia Comput. Sci.*, vol. 120, pp. 602–607, 2017.
- [35] M. I. Ibrahimy, M. R. Ahsan, and O. O. Khalifa, "Design and optimization of levenberg-marquardt based neural network classifier for EMG signals to identify hand motions," *Meas. Sci. Rev.*, vol. 13, no. 3, pp. 142–151, 2013.
- [36] H. Ullah et al., "Levenberg-Marquardt Backpropagation for Numerical Treatment of Micropolar Flow in a Porous Channel with Mass Injection," *Complexity*, vol. 2021, 2021.
- [37] P. Malik, A. Gehlot, R. Singh, L. R. Gupta, and A. K. Thakur, "A Review on ANN Based Model for Solar Radiation and Wind Speed Prediction with Real-Time Data," *Arch. Comput. Methods Eng.*, no. January, 2022.

# Expanding Louvain Algorithm for Clustering Relationship Formation

Murniyati<sup>1</sup>, Achmad Benny Mutiara<sup>2</sup>, Setia Wirawan<sup>3</sup>, Tristyanti Yusnitasari<sup>4</sup>, Dyah Anggraini<sup>5</sup>  
Computer Science and Information Technology of Gunadarma University, Jakarta, Indonesia<sup>1, 2, 3, 4, 5</sup>

**Abstract**—Community detection is a method to determine and to discover the existence of cluster or group that share the same interest, hobbies, purposes, projects, lifestyles, location or profession. There are some example of community detection algorithms that have been developed, such as strongly connected components algorithm, weakly connected components, label propagation, triangle count and average clustering coefficient, spectral optimization, Newman and Louvain modularity algorithm. Louvain method is the most efficient algorithm to detect communities in large scale network. Expansion of the Louvain Algorithm is carried out by forming a community based on connections between nodes (users) which are developed by adding weights to nodes to form clusters or referred to as clustering relationships. The next step is to perform weighting based on user relationships using a weighting algorithm that is formed by considering user account activity, such as giving each other recommendation comments, or to decide whether the relationship between the followers and the following is exist or not. The results of this study are the best modularity created with a value of 0.879 and the cluster test is 0.776.

**Keywords**—Community detection; Louvain algorithm; modularity; network clustering relationship

## I. INTRODUCTION

Social networks become a well-known instrument to disseminate information and to connect people who have the same thoughts. Public accessibility of this network with the capability to share opinions, thoughts, information and experience, offer tremendous promises for companies and governments. Apart from individuals who use this network to connect with friends or relatives, many companies and governments start to leverage social network platform to deliver their services to the society, citizen and client. Trust is a crucial issues of an effective social network. A social network is defined as a graph that contain set of nodes. Nodes represent objects, actors, people, or organizations while links express collaborations, communications or interactions. A complex network shows a very dense (such as a network of friendships or collaborator) or sparse network[1] [2][3][4].

Social network has received much attention and has been studied during the last decades including community detection in large and complex network[5] The aim of community detection is to divide the network and to illustrate this network in graph's form. The nodes (objects, actors, people, or organizations) that have relationship or correlations are said to be in the same community. Community can be used to obtain various purposes (such as finding targets who like similar product, encounter target markets, defining product ratings popularity, determining product recommendations and much

more [6]. There are lots of information on social media that can be used to cluster the data based on their similarity. Individuals can find peoples' biodata or what they share in their social media. This behavior definitely influenced the consequences of user curiosity. This research will form communities based on books at Gramedia Pustaka Utama.

Clustering is the assignment to separate the object inside the population into number of groups based on certain characteristic. Community will be detected based on the obtained information. Unfortunately, there are some drawback on the previous community detection algorithm. The previous algorithm could not perform well on large data set. Most studies, maximize the quality function to determine the community. This method known as modularity. In modularity, the nodes inside the same community are highly connected but loosely linked to nodes outside their community[7]. Modularity maximization calculates the quality of a particular clustering of a network into communities. It is a matter of processing speed and most of its algorithm use heuristic technique. Amongst the most efficient modularity maximization algorithm is using Louvain method. Louvain methods is a method to extract communities form large scale network. The results obtained from Louvain method give good modularity quality which describe the closeness relationship between nodes. This closeness relationship will be used to compute the confidence level of users in the recommendation system that will be developed[8][9]. Throughout developing Louvain algorithm, the usage of large amount data set from social media network is capable and will generate fast clustering relationship.

Social network influenced the behaviour of users. Therefore, in this research a recommendation system was developed based on the relations that were obtained from the social network[10]. The necessary step to implement clustering relationship formation was by performing scrapping. Scrapping was achieved to obtain relation on Gramedia Pustaka Utama account. The steps taken are scrapping first to get existing relationships with Instagram accounts, then doing weights and forming clusters by developing the Louvain algorithm. The purpose of this study is to develop the Louvain algorithm to produce clustering relationships so that you can see your circle of friends on Instagram social media. The results of this study can be used to produce a recommendation system that is not only based on ratings and reviews, but based on clustering relationships. Purity is used to measure the success of the clusters formed.

## II. LITERATURE

### A. Community Detection Algorithm

Detecting communities is a very essential task as communities helps us in grouping users showing similar behaviour and in this way the social network can be divided into different clusters of nodes with same behaviour. This community information can help us take useful decisions and extract important information about users in a particular community [11].

Community detection is an algorithm to create groups or partitions and to evaluate the formations of these groups. There are many community detection algorithm propose by different researchers, such as :

- Strongly connected components (SCC) are one of the initial graphing algorithms. This algorithm was described by Tarjan in 1972. SCC algorithm defined directed graph into strongly connected components which is a classic application of depth-first reach algorithm [12][13].
- Weakly Connected Components or Union Find algorithm found set of connected nodes in undirected graph, where each node is reachable from any other nodes in the same set. Weakly Connected Components is different from SCC. This algorithm only require an existing path between a pair of nodes in one direction, whereas SCC algorithm requires path to exist in both direction. Like SCC, Weakly Connected Component algorithm often used as initial stage in analyzing the graph's structure [13].
- Label Propagation Algorithm is a fast algorithm to discover communities in graphs. To detect a community using the structure of the network does not require any predefined objective functions nor prior information regarding the community. One of the interesting features of Label Propagation Algorithm is that the algorithm has the option to control the initial label to narrow down the obtained solution. At initial stage, each node has a unique label. Afterwards, labels are assigned iteratively to nodes in a random sequential order in such a way that nodes take the most frequent label of its neighborhood. The label relocation unites when there is no more alteration in the node label. Groups of nodes that have identical nodes at convergence form communities. Although this approach is efficient and does not require user-defined parameter, it is not deterministic due to the random choice of nodes to be labeled and the possibly large number of edges explored during the iterations [14].
- Triangle count and average clustering coefficient algorithm computes the number of triangles in the graph. Triangle is three set of nodes where each node is connected to the other two. In graph terminology triangle is known as 3-clique. The Triangle Count algorithm in Graph Data Science (GDS) library discover triangles only in undirected graphs. Triangle count has gained popularity in Social Network Analysis (SNA). This algorithm is used to detect communities and to measure

the cohesiveness of these communities. It also can be used to define the constancy of a network. Moreover, triangle count can be employed to compute the networks' indexes, such as clustering coefficient and local grouping coefficient [15].

- Another method for community detection is a spectral approach by Newman, which is a top-down hierarchical one that depends on eigenvectors of the modularity matrix. This approach works by iteratively separating the network into two components so that the modularity is maximized.[16]
- Another commonly used method for community detection is based on the modularity maximization , which calculates the quality of a particular clustering of a network into communities. The intuition behind modularity is that nodes inside the same community are highly connected but loosely linked to nodes outside their community.[17][18]
- Louvain modularity is an algorithm for detecting communities in a network. This algorithm maximizes modularity value of each community. Modularity quantifies the quality of assigning nodes to the community by evaluating how much more tightly connected the nodes in the community are, compared to how connected they would be in a random network. Louvain algorithm is one of the fastest modularity algorithm that perform well on large scale graph. This algorithm reveals the hierarchy of communities at different scales, which is useful for understanding the global functioning of the network. To understand Louvain's modularity algorithm, it is important to learn modularity in general.

It can be said that community detection is an effort and process to determine and find a group of people who have the same or the same interests. Community detection or clustering graph becomes part of data analysis in various fields; computer science, science, social network analysis and internet applications. As the data grows on exploratory power. Community detection is widely used in graph analysis. Given the graph  $G = (V, E)$ , the goal of the community detection problem is to identify the partitioning of nodes into "communities" (or "clusters") so that related nodes are assigned to the same community and different or unrelated nodes are assigned to the same community. Different. The community detection problem differs from the classic graph partition problem in that neither the number of communities nor their size distribution are known a priori. Due to its ability to uncover structurally coherent node modules, community detection has become a structure discovery tool in a number of scientific and industrial applications, including biological sciences, social networking, retail and finance.

The concept of community detection exists in network science as a method for finding communities in complex systems through graphical representations. The community detection method finds subnetworks statistically between nodes or graphs in the same community rather than nodes in different communities [19].

The core of community detection is the idea of modularity, a metric of which differs below:

$$Q = \frac{1}{2m} \sum_{i,j} A_{ij} \left( A_{ij} - \frac{k_i k_j}{2} \right) \delta(g_i, g_j) \quad (1)$$

In the above difference Q is modularity,  $A_{ij}$  is the edge weight between vertices  $i$  and  $j$ ,  $k_i$  is the total weight of all edges connecting node  $i$  with all other vertices, and  $2m$  is the total weight of all edges in the graph. The Kronecker delta function  $\delta(g_i, g_j)$  will evaluate to one if the nodes  $i$  and  $j$  belong to the same group, and are zero. Modularity is the defining state of how one decides to divide the network. An unshared network is a network in which each node in its own community will have a modularity equal to zero. The goal of community detection is to find a community that can maximize modularity. There are many efficient algorithms to maximize modularity, including spectral clustering [16] [20]. Fig. 1 shows multiple networks with maximum increased modularity. Notice how the community structure becomes clearer as the value of modularity increases.

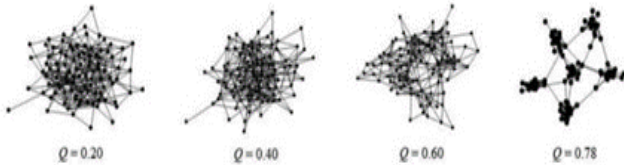


Fig. 1. Networks with different maximum modularity (Blondel et al., 2008)

### B. Louvain Algorithm

Louvain's algorithm shows an algorithm that directly maximizes modularity with 2 phase algorithm. This first algorithm consists of nodes moving one by one in one of the neighboring communities to get the maximum increase in modularity, the nodes can be moved multiple times and this procedure stops if maximum locales are obtained, that is, when there is no more movement which increases the modularity. The second algorithm is the formation of a Meta graph where the nodes are the communities found in phase 1 and the links represent the number of connections between communities. The Louvain algorithm is an unsupervised algorithm that does not require input on the number of communities or size before running. The Louvain algorithm is divided into 2 phases, namely Optimizing Modularity and Community Aggregation.

Louvain's algorithm is one of many algorithms for community detection. One of the advantages of the Louvain Algorithm is that it detects communities with maximum modularity and is also faster than other algorithms.

Louvain's algorithm was first introduced to find the Newman-Girvan high partition modularity.

$$M = \frac{1}{2m} \sum_{i,j} A_{ij} (A_{ij} - p_{ij}) \delta(c_i, c_j) = \frac{1}{2m} \sum_{i,j} A_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \quad (2)$$

$A_{ij}$  is the neighbor matrix entry which represents the weight of the edge connecting vertices  $i$  and  $j$ ,  $k_i = \sum_j A_{ij}$  is the degree of the node  $i$ ,  $c_i$  is the community,  $\delta(c_i, c_j)$  value 1 if  $c_i = c_j$  and 0 if otherwise.  $m = \frac{1}{2} \sum_{i,j} A_{ij}$  is the sum of the weight of all sides on the graph.

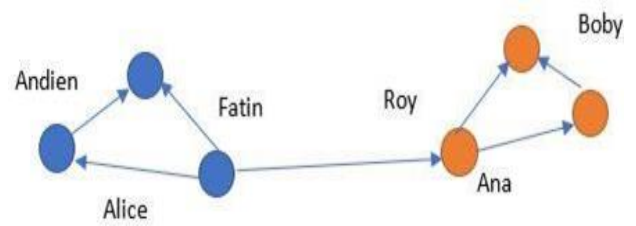


Fig. 2. Illustration of a community formed with louvain's algorithm

Louvain's algorithm in Fig. 2. finds two communities with three members on each community. Andien, Alice and Fatin are friends with each other, as are Roy, Ana and Bobby. Roy is the only one who has friends in both communities, but Roy has more friends who have the same characteristics in community two therefore Roy is in that community.

### C. Modularity

Modularity is a measure of how well a group has been partitioned into clusters. It compares the relationships in a cluster against what is expected for a random number of connections. Criteria is known as modularity, its definition involves a comparison of the number of in-cluster links in a real network and the expected number of links in a random graph (regardless of community structure)[7].

## III. LITERATURE

The steps taken to form clustering relationships are as shown in Fig. 3.

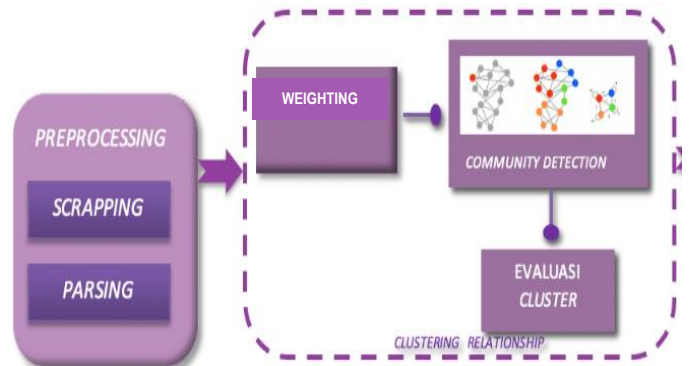


Fig. 3. System architecture

In Fig. 4, the system architecture begins with collecting data sets from Gramedia Pustaka Utama. The information obtained from Gramedia Pustaka Utama was related to the object of research, namely bookstores and novels.

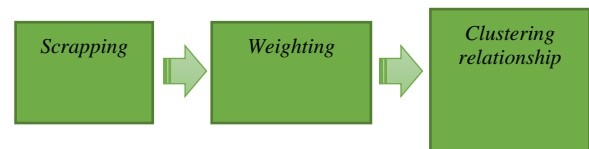


Fig. 4. Stages of forming a clustering relationship

After collecting the data set, the preprocessing stage is carried out using a social-based concept to take into account the relationship between users and other users in a social networking application service. The stage that is done is by

doing scrapping to get connectedness between users and aims to see how much users trust other users (shown in Fig. 5).



Fig. 5. Scrapping stage

Network connectivity on Gramedia Pustaka Utama was given a weight and then community detection is carried out using the Louvain method. The next step is to build a trust matrix which is represented by the results of weighting and relationship. This procedure is called a clustering relationship. The next stage is to compute the level of similarity of a user. In this stage every user will produce a rating prediction using collaborative filtering based similarity

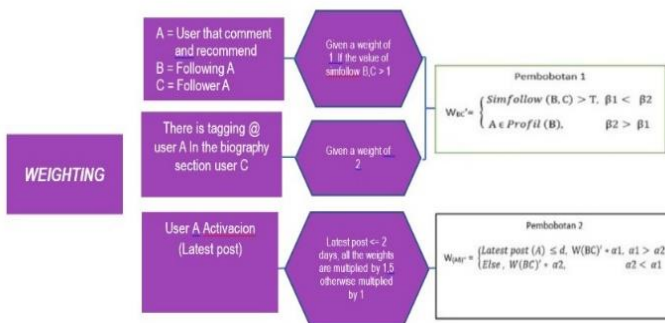


Fig. 6. Weighting process

The weighting procedure is a process that was performed to compute the trust value between users. The process was expected to obtain cluster's weight or value when creating a cluster relationship. This weight or relation value can be used in forming matrix trust. The weight is based on the relationship between A, B, C, users who recommend others and the activeness of user.

The weighting was performed in 2 stages. The first stage was computing the closeness of relationship using variable 1 as WBC (WBC weight). The given value for variable  $\beta_1$  is 1 and for variable  $\beta_2$  is 2. The second stage of the weighting process is to see user's activity. This step is require to obtain the users that are active to form a cluster. The variable that used for 2 is WAB (WAB weight). The given value for variable  $\alpha_1$  is 1.5 and for variable  $\alpha_2$  is 1 as shown in Fig. 6.

In algorithm, weight 1 is the weight for user A who has proximity (simfollow) between B and C > the threshold value, then it is being given weight of 1 and if in Biography C there is a tagging @user A then it is given weight of 2. The value of weigh 1 is set to variable  $\beta_1$ , the weight value of 2 is set to variable  $\beta_2$  where  $\beta_1 < \beta_2$ .

The next step is performed by increasing the weight level. For the second weighting, the weight can be given by looking at user activity (User A) from user A's last post. If user A's last post is less than 2 days, then all user A's weights are multiplied by 1.5, if not multiplied only by 1.

### Algorithm 1: Weighting Algorithm

Input : User A

Output: Followers A(C)

Following A (B)

1. Get Follower A
2. Get Following B
3. Give a weight of 1, if the followers intersect with the following, and give a weight of 2 if the Follower C user has a user A tag in the biography section, using a weighting formula 1:

$$W_{BC} = \begin{cases} Simfollow (B,C) > T, \beta_1 < \beta_2 \\ A \in Profil (B), \beta_1 > \beta_2 \end{cases}$$

4. If the latest post is less than or equal to 2, the weight is multiplied by 1.5 for users connected to the A besides that the weight is multiplied by 1, using the weighting formula 2 :

$$W_{AB} = \begin{cases} Latest Post (A) \leq d, W(BC)' * \alpha_1, \alpha_1 > \alpha_2 \\ Else, W (BC)' * \alpha_2, \alpha_2 < \alpha_1 \end{cases}$$

The multiplier of 1.5 is assigned a variable ( $\alpha_1$ ) and the multiplier of 1 is assigned a variable ( $\alpha_2$ ). The value of  $\alpha_1 < \alpha_2$ . This function is to indicate the activeness of the user. For example, user Osterdamm (A) comments on the book 'The Miracle of Mindbody Medicine\_New', (B) is a user in Following Osterdamm, (C) is all Follower Osterdamm users if they are related or are friends then they are given a weight of 1, if Follower Osterdamm in his biography has the tag '@Osterdamm' then it is given a weight of 2, if Follower Osterdamm last posted less than 2 days then the weight is multiplied by 1.5.

## IV. RESULT

Community detection is a method to find communities in a large and complex network. This method optimized modularity. There are many algorithm to maximize modularity. For example spectral clustering (Newman, 2006) and fast unfolding (Blondel et al., 2008). This research propose Louvain algorithm to detect communities. Louvain algorithm is considered the most suitable method to detect the communities since the algorithm works well on large and complex network. This algorithm does not require data input (Unsupervised learning). Moreover the algorithm can form clusters/communities faster compare to other algorithms. Louvain algorithm is the development of the existing community detection algorithms. Louvain is unsupervised learning algorithm. It does not require input of communities' size or number. The algorithm is divided into two phase; Modularity optimization and community aggregation. Modularity is used in this research to measure how well a group has partitioned into clusters.

Below is the equation to compute modularity:

$$M = \frac{1}{2m} \sum_{i,j} (A_{ij} - p_{ij}) \delta(c_i, c_j) = \frac{1}{2m} \sum_{i,j} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j) \quad (3)$$

$A_{ij}$  is the adjacency matrix that represent the edges weight which connect node  $i$  and node  $j$ ,  $k_i = \sum_j A_{ij}$ .  $A_{ij}$  is the degree of node  $i$ ,  $c_i$  is its community,  $\delta$ -function  $\delta(u, v) = 1$  if  $u = v$  and 0 otherwise.  $m = \frac{1}{2} \sum_{ij} A_{ij}$  is the sum of all weight in a graph.

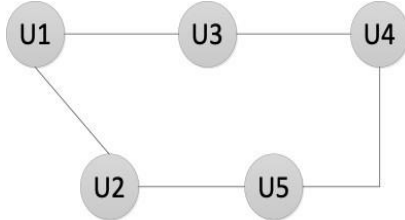


Fig. 7. A simple graph

Fig. 6 shows a graph with 5 nodes (U1, U2, U3, U4 and U5) and 5 edges (U1-U3, U3-U4, U4-U5, U5-U2, U2-U1). According to the definition of Neighbor Matrix explained by  $A_{ij}$  in the previous section, it will look like Table I.

TABLE I. TABLE GRAPH REPRESENTATION IN MATRIX

	U1	U2	U3	U4	U5
U1	0	1	1	0	0
U2	1	0	0	0	1
U3	1	0	0	1	0
U4	0	0	1	0	1
U5	0	1	0	1	0

Considering assigning the partition as:

Partition 1, U1-U1, U3, U4

Partition 2, U2-U2, U5

Using Modularity equation (1)

1) Node U1 to U1

$$Q = 1/2 * 5((0 - 2 * 2)/(2 * 5)) * 1$$

#node U 1 – U 1

Same member (one partition)

2) Node U1 to U2

$$Q = 1/2 * 5((1 - 2 * 2)/(2 * 5)) * 0$$

#node U 1 – U 2

Not the same member (not one partition)

3) Node U1 to U4

$$Q = 1/2 * 5((0 - 2 * 2)/(2 * 5)) * 1$$

#node U 1 – U 4

Same member (one partition)

4) Node U1 to U5

$$Q = 1/2 * 5((0 - 2 * 2)/(2 * 5)) * 0 \text{ #node U 1 – U 5}$$

Not the same member (not one partition)

5) Node U1 to U3

$$Q = 1/2 * 5((1 - 2 * 2)/(2 * 5)) * 1$$

#node U 1 – U 3

Same member (one partition)

Modularity optimization in this research is illustrated as follows.  $\sum_{in}$  is the sum of edges' weight in  $C$ ,  $\sum_{tot}$  is the sum of all edges to node in  $C$ ,  $k_i$  is the sum of weight from all edges in node  $i$ ,  $k_i, in$  is the sum of edges from node  $i$  to node inside community  $C$  and  $m$  is the sum of weights from all edges in the graph.

Louvain Algorithm PASS 1 and Louvain Algorithm PASS 2. The first step in Louvain Pass 1 algorithm is to select the initial node and computing the modularity transformation that might occur when a node join and formed community with this node close neighbor. The next step is that initial node join the node with the highest modularity change. This process will be repeated for each node until the community is formed. Communities are combined to create super communities and relationships between this super node counts as the sum of the previous links (Self-loop represents the previous relationship and is now hidden in the super node) The resulting graph using the Louvain PASS 1 algorithm can be seen in Fig. 7.

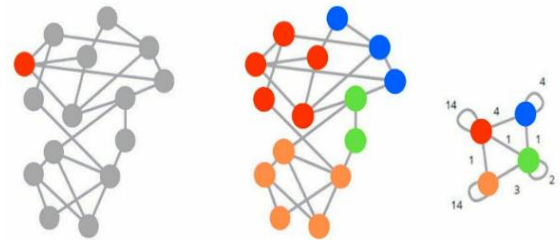


Fig. 8. The created graph with PASS 1 algorithm

The steps for creating a graph in the Louvain PASS 1 algorithm are as follows:

**Algorithm 2: LOUVAIN PASS 1**

Require :  $G = (V, E, w)$  a weighted graph

Ensure : a partition  $P$  of  $V$

Local : increase  $\leftarrow$  true

Local :  $P$  current partition of  $V$

Begin

Forall the nodes  $I$  do

$P[i] \leftarrow \{i\}$

INIT ( $i$ )

While increase do

increase  $\leftarrow$  false

forall the nodes  $i$  do

Remove ( $i, C_{old}$ )

$C \leftarrow P[j] \mid (i, j \in E) \cup \{C_{old}\}$

$C_{old} \leftarrow P[i]$

$C_{new} \leftarrow \operatorname{argmax}_{c \in c} \{GAIN(i, C)\}$

INSERT ( $I, C_{new}$ )

if  $C_{old} \neq C_{new}$  then

increase  $\leftarrow$  true

The resulting graph using the Louvain PASS 2 algorithm can be seen in Fig. 8 below:

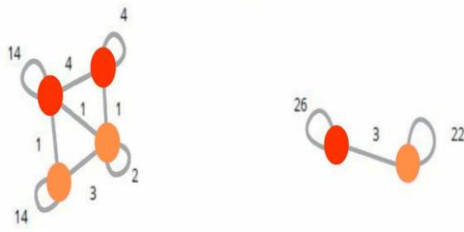


Fig. 9. The created graph with PASS 2 algorithm

The steps for creating a graph in the Louvain PASS 2 algorithm are as follows:

**Algorithm 3: LOUVAIN PASS 2**

Require :  $G = (V,E,w)$  a weighted graph

Ensure : a partition  $P$  of  $V$

Begin

Repeat

$P \leftarrow \text{ONEPASS}(G)$

$G \leftarrow \text{Partition} - \text{to} - \text{Graph}(P,G)$

until no improvement is possible

In the Louvain PASS 2 algorithm, steps 1 and 2 are repeated in the path until no further growth in modularity. The PASS 2 algorithm will also repeated until the number of iterations has occurred.

In the initial research conducted using the Louvain PASS 1 and PASS 2 algorithms, the resulting graph can be seen in Fig. 9 and Fig. 10.

Modularity for best partition: 0.9568860789869055

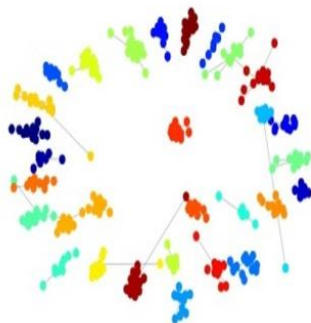


Fig. 10. Visualization of clustering relationship using the louvain pass 1 algorithm

Fig. 9 shows the obtained graph using Louvain algorithm. This figure created from some of the data (219570 data from Gramedia Pustaka Utama’s users). The visualization of Louvain PASS 2 algorithm can be seen in Fig. 10. The obtained value of modularity is 0.5, indicating that the community detection was not good enough.

The proposed Louvain algorithm. In the previous Louvain Algorithm, there are 2 phases of modularity formation. The developed algorithm starts from the input which is the result of weighting the nodes to form communities. The formation starting by defining different communities of each network node. In the initial partition, the number of communities is equal to the number of nodes.

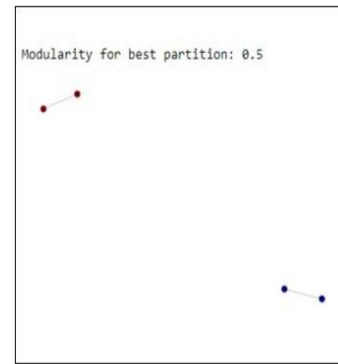


Fig. 11. Visualization of clustering relationship using the louvain pass 2 algorithm

**Algorithm 4: LOUVAIN MURNI**

Input : Dataset of relations between Instagram users

Output : Clustering Relation Graph, list of cluster member

1. Random user to be used as a start point (make  $S$  variable)
  2. Remove users who are neighbors with  $S$  (save to variable  $N$ ) from cluster 1 to a new cluster (cluster 2)
  3. Compute modularity (make it  $mod\_new$  variable); using the modularity formula:
- $$M = \frac{1}{2m} \sum_{i,j} (A_{ij} - p_{ij}) \delta(c_i, c_j) = \frac{1}{2m} \sum_{i,j} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j)$$
4. If  $mod\_new$  is greater than  $mod$ , make  $X=X+1$  and set  $mod = mod\_new$ , if smaller than return to cluster 1
  5. Set  $S = N$
  6. Repeat step (4) ,until all users are counted

For each node  $i$ , the neighboring nodes  $i$ , namely node  $j$  will be considered. Then the value of obtained modularity will be evaluated by removing node  $i$  from its community and then placing node  $i$  in the community of node  $j$ . Node  $i$ , then positioned in the community that provides the greatest value of profit, but only if the value of the gain is positive. If there is no possibility of a positive profit value, node  $i$  will remain in the community from which the node originated. This process will be repeated successively for all nodes until no further improvement can be achieved. With the fulfillment of these conditions, the first phase of the algorithm in this study has been completed.

The previous Louvain algorithm which consists of 2 Louvain phases produces a super cluster which is only a few large clusters. Pure Louvain Algorithm which was developed to get user clusters that have relationships obtained from the previous weighting process, this is needed for the stage of forming a recommendation system based on Trusted Friend.

The results of the pure Louvain Algorithm will be represented in the form of a graph, a list of cluster members and the relationships between users

V. PURITY APPROACH

Purity or a measure of purity historically was the first measurement used in the context of community detection used by Girvan and Newman in their article. Purity has gone by a variety of different names in several articles making it difficult to name a complete list. The purity of a part relative to part Y is expressed in the following equation:

$$Pur(x, Y) = \max \frac{nij}{ni+} \quad (4)$$

In other words, the first thing to do is to identify the part with the largest intersection and then calculate the proportion of the elements. The greater the intersection and the greater the purity value, the greater the correspondence between the two parts being analyzed. Then the total partition X relative to the partition is obtained by adding up the purity of each xi, then given a weight using the following equation:

$$Pur(X, Y) = \sum_i \frac{ni+}{n} Pur(xi, Y) \quad (5)$$

The upper limit on purity is 1, which corresponds to a perfect match between each partition, while the lower limit is 0 which is the opposite value of the upper limit. Purity is not a symmetrical measure, meaning that in the process, purity is relative to the amount considered in each part. Therefore, in general PUR(X,Y) is not the same as PUR(Y,X).

From a community detection point of view, two different purity measurements can be used, depending on whether to calculate the estimated community purity relative to the true value, or vice versa. In cluster analysis, the first version is generally used, and is called simply Purity, while the second version is Inverse Purity. It is difficult to determine which one is actually used in the case of existing community detection. Girvan and Newman provide a very concise description of the size being processed. Purity tends to favor algorithms that identify many small communities. In the most extreme case, if the algorithm identifies n communities containing one node each, one of the clusters gets the maximum purity, because each estimated community is perfectly pure. In contrast, reverse purity supports algorithms that detect multiple large communities. The most extreme case occurs when the algorithm places all nodes in the same community, then a cluster gets the maximum purity, because each community is actually perfectly pure: all the nodes it belongs to belong to the same estimated community. To solve this problem, Newman introduces an additional solution: when the estimated community is majority in some actual community, all the nodes in question are considered to be the wrong classification. The solution generally adopted in cluster analysis consists mainly of processing the F-Measure, which is the average of the harmonics of the two purity versions:

$$F(X, Y) = \frac{2.Pur(X*Y*Pur(Y,X))}{Pur(X,Y)+Pur(Y,X)} \quad (6)$$

The measure obtained from the above equation is symmetrical, and this combination is expected to resolve the aforementioned bias. This approach provides a solution in a similar way by underestimating and overestimating the number of communities. The purity value is calculated for all clusters formed, namely 27 clusters. Calculating the purity value,

starting from the cluster that has a majority value to users who tag the preferred book, as shown in Table II below.

TABLE II. THE OVERALL PURITY OF THE CLUSTER

	Lower Bound	Maximum	Purity
Cluster 1 until Cluster 27	0.0	1	0.776

For example, Cluster 1 the majority value is 25, Each cluster is calculated as the majority value, so that the total of the entire cluster has a purity value of 0.776 as shown in Fig. 11:

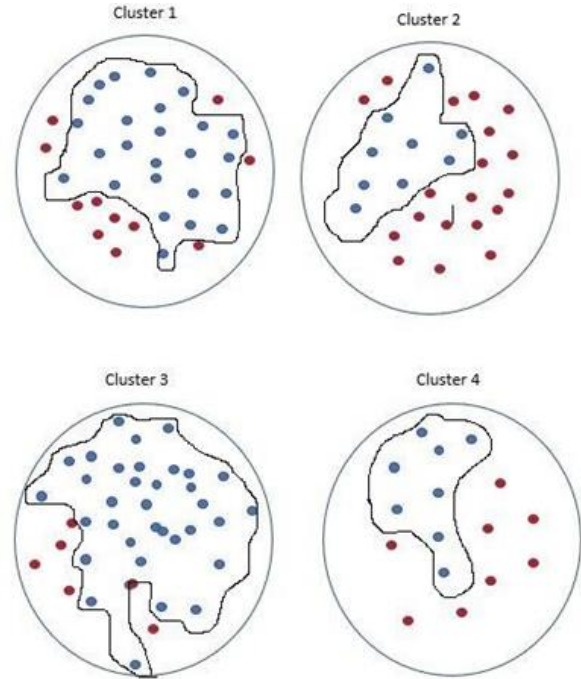


Fig. 12. Purity as an external evaluation criterion for the quality of cluster 1 to cluster 27.

VI. CONCLUSION

Based on a number of tests and analysis of the results of this study, it can be concluded that:

- 1) The social network on Gramedia Pustaka Utama account forms a clustering relationship that is used for the recommendation system by weighting and community detection. The weighted value (α, ) given is proven to affect the results of the community that is formed.
- 2) Clustering Relationship succeeded in forming clusters using the Louvain algorithm, as many as 27 clusters with the best value of high modularity, namely 0.879.
- 3) The value of the modularity of the community that is formed is influenced by the number of relationships between community members where the denser the relationships in the community, the value of modularity will increase or be higher.
- 4) The use of algorithms with modularity optimization has slightly better results because modularity shows how well the community on the network is.



5) Evaluation of the cluster formed using purity produces a satisfactory value of 0.776.

#### ACKNOWLEDGMENT

The authors would like to acknowledge to Research Center of Gunadarma University and Gramedia Pustaka Utama. The authors also thank to colleagues for input and support.

#### REFERENCES

- [1] W. Pedrycz and S.-M. Chen, Eds., *Social Networks: A Framework of Computational Intelligence*, vol. 526. Cham: Springer International Publishing, 2014.
- [2] Murniyati, A. B. Mutiara, S. Wirawan, and T. Yusnitasari, "Social Networks for Establishing Clustering Relationships on Instagram Social Media," *I T A L I E N I S C H*, vol. 11, no. 2, pp. 394–407, 2021, doi: 10.1115/italienisch.v11i2.132.
- [3] S. Lobanova and A. Chepovskiy, "Combined method to detect communities in graphs of interacting objects," no. 4 (42), pp. 64–73, 2017. [Online]. Available: <https://bijournal.hse.ru/en/2017--4>.
- [4] L. Ma and Y. Zhang, "Hierarchical social network analysis using multi-agent systems: A school system case," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2014, pp. 1412–1419, doi: 10.1109/SMC.2014.6974113.
- [5] Y. W. and Q. Y. Y. Zhang, "Community Discovery in Twitter Based on User Interests," *J. Comput. Inf. Syst.*, vol. 8 N0. 3, pp. 991–1000, 2012.
- [6] W. M. and A. A. C. N. Utami, "Analisis dan Implementasi Community Detection Menggunakan Algoritma Girvan and Newman Dalam Sosial Network," 2013.
- [7] Newman, M.E.J., "Modularity and community structure in networks," *P. Natl. Acad. Sci. USA*, vol. 103, pp. 8577–8582, 2008.
- [8] A. Stachowiak, "Propagating and Aggregating Trust with Uncertainty Measure," 2011.
- [9] A. Stachowiak, "Uncertainty-Preserving Trust Prediction in Social Networks," 2014.
- [10] F. E. Walter, S. Battiston, and F. Schweitzer, "A model of a trust-based recommendation system on a social network," *Auton. Agent. Multi. Agent. Syst.*, vol. 16, no. 1, pp. 57–74, 2008, doi: 10.1007/s10458-007-9021-x.
- [11] P. Kumar, R. Jain, S. Chaudhary, and S. Kumar, "Solving Community Detection in Social Networks: A comprehensive study," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 239–345, doi: 10.1109/ICCMC51019.2021.9418412.
- [12] S. Raghavan, "Twinless Strongly Connected Components," in *Perspectives in Operations Research: Papers in Honor of Saul Gass' 80th Birthday*, F. B. Alt, M. C. Fu, and B. L. Golden, Eds. Boston, MA: Springer US, 2006, pp. 285–304.
- [13] I. T. Neo4J, "Neo4j Graph Data Science Library," 2022. <https://neo4j.com/docs/graph-data-science/current/algorithms/louvain/>.
- [14] B. Laassem, A. Idarrou, L. Boujlaleb, and M. Iggane, "Label propagation algorithm for community detection based on Coulomb's law," *Phys. A Stat. Mech. its Appl.*, vol. 593, p. 126881, 2022, doi: <https://doi.org/10.1016/j.physa.2022.126881>.
- [15] S. Putrevu, "A Note on Detection of Communities in Social Networks," 2020, doi: 10.18535/ijecs/v9i03.4449.
- [16] B. A.-L. and W. D. J. Newman M E J, "The Structure and Dynamics of Networks," *Princet. Univ. Press. Princet.*, 2006.
- [17] M. Newman, MEJ and Girvan, "Finding and evaluating community structure in networks," 2004, doi: 10.1103/Phys-RevE.69.026113.
- [18] J. Yang, J. J. McAuley, and J. Leskovec, "Community Detection in Networks with Node Attributes," *CoRR*, vol. abs/1401.7267, 2014, [Online]. Available: <http://arxiv.org/abs/1401.7267>.
- [19] S. Zhang, Z. Liu, and W. Dou, "An Enhanced Community Detection Method Based on Neighborhood Similarity," in *Proceedings of the 2012 Second International Conference on Cloud and Green Computing*, 2012, pp. 493–500, doi: 10.1109/CGC.2012.71.
- [20] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech. Theory Exp.*, vol. 2008, no. 10, p. P10008, Oct. 2008, doi: 10.1088/1742-5468/2008/10/P10008.

# Implementation Failure Recovery Mechanism using VLAN ID in Software Defined Networks

Heru Nurwarsito<sup>1</sup>, Galih Prasetyo<sup>2</sup>

Faculty of Computer Science, University of Brawijaya, Malang, Indonesia<sup>1,2</sup>

**Abstract**—Link failure is a common problem that occurs in software-defined networks. The most proposed approach for failure recovery is to use pre-configured backup paths in the switch. However, it may increase the number of traffic packets after the traffic is rerouted through the backup path. In this research, the proposed method is the implementation of a failure recovery mechanism by utilizing the fast failover group feature in OpenFlow to store pre-configured backup paths in the switch. The disrupted traffic packets will be labeled using the VLAN ID, which can be used as a matching field. Due to this capability, VLAN ID can aggregate traffic packets into one entry table as a match field in the forwarding rules. Through implementation and evaluation, it is shown that the system can build a backup path in the switch and reroute the disrupted traffic to the backup path. Based on the parameters used, the results show that the proposed approach achieves a recovery time of around 1.02-1.26ms. Additionally, it can reduce the number of traffic packets and has a low amount of packet loss compared to previous methods.

**Keywords**—Software-defined networks; openflow; link failure; failure recovery; VLAN ID; fast failover

## I. INTRODUCTION

Software-defined networking is a new paradigm that changes the current network infrastructure. The SDN concept is to break down the network infrastructure by combining the control logic (control plane) of routers and switches that forward traffic (data plane) [1]. Unlike the conventional network concept, where the control plane and forwarding are tied directly to the same network device. This causes the network administrator to have to configure the device manually. Another disadvantage of conventional networks is that when a device encounters a problem, the network administrator must fix the problem directly on the device.

In SDN, several problems can occur when the routing process is executed, one of which is a link failure. Link failure is one of the problems on the network that causes delays and even packet loss so the throughput value decreases. Link failures consist of direct or indirect failures. In the case of direct failures, the switches detect the failure immediately and recover quickly, whereas, in the case of indirect failures, the link failure is not detected by the respective switch despite traffic overhead. Unidirectional link failures disrupt traffic and create a loop in the switch topology. Multiple link failures reduce network reliability performance [2].

Based on previous research that used proactive SDN methods to solve the link failure problem. When a link fails, a predetermined backup path is created and used in this proactive

method. The switch closest to the link failure point will then reroute through the backup path to reach the destination [3]. Then in another research, using the rerouting method to overcome link failure was examined. When a link fails, high-priority packets are temporarily diverted to an alternate path, and then the packet is sent to the destination using the rerouting method to find the shortest path to the destination [4].

In the several methods previously mentioned, these methods can overcome link failures without involving the controller, thereby reducing failover time on SDN. However, these methods have several downsides, such as the need for a large number of flow entries to build backup lines and complex processes to maintain the lines alive. The backup path that has been created is then difficult to modify or adapt to changing network conditions, allowing for the possibility of congestion during the rerouting procedure.

Based on the problems described above and previous research, a test simulation will be developed to implement failure recovery on SDN by incorporating the fast failover group and VLAN ID features. This method works by creating a backup path for each link using the SDN architecture's fast failover group feature. Next, enable the VLAN ID feature, which is used as a match field in forwarding rules, to reduce traffic when packets are diverted to the backup path. This approach allows the system to recover from failures without involving the controller, reducing recovery time and minimizing the use of flow table memory on switches during link failures.

The rest of the paper is organized as follows: Section II provides a brief review of the literature relevant to our work and describes the main concepts around our approach. Section III explains the methodology of our approach is presented. Section IV explains the results and findings of the research. Section V is the discussion that presents the comparison result and findings with the previous studies. In Section VI, the conclusion of this research.

## II. RELATED WORK

### A. Background

The Group Table-based Rerouting (GTR) method is one of the approaches used to find responses to single link failures through the fast failover (FF) group [5]. In general, a backup path is created for each link between the source and the destination; however, the backup path proposed in this research is created by calculating the most efficient path between adjacent switches. The controller periodically updates the lookup table on the controller and the FF group table on the

switch to determine what changes have occurred in the topology and network traffic. The research proposed a protection scheme for source routing-based backup links [6]. The proposed method controls packets going to the backup link by updating the source routing header. With SDN source routing, it can use VLAN tags to store packet routes in the source routing headers instead of being stored in switch flow entries. By using this method, it can reduce the number of flow entries needed to build a backup link. However, these approaches have limitations in the implementation process, which is quite complex to overcome existing problems in the fast failover mechanism.

Furthermore, in the research conducted [7], flow-based network management was proposed that can be programmed on the OpenFlow network. Researchers propose a method called Path Monitoring (PathMon), which encodes flow and path information as tags that can work flexibly. This method is implemented on a switch and uses OpenFlow 1.3, which supports VLAN tags to encode flow and path information as flexible tags so that the statistical data obtained to monitor network traffic on OpenFlow is more specific. The purpose of using VLAN IDs in this research is to make it easier for network administrators to monitor network traffic and collect the different statistical information needed. In further research, there are problems with the data center caused by the detection of elephant flow, which resulted in high network latency [8]. The proposed method is to use multipath routing, which can break down elephant flows into several mice flows that are distributed evenly on the network without detecting elephant flows. In this research, VLAN-based routing is used to reduce flow entry consumption, where the controller can instruct the switch to enter the path ID in the VLAN ID field in each packet from the flow when routing to the switch. The results showed a 32% reduction in flow entry on switches compared to the method without using VLANs. This research is distinct from one another in that researchers raise different issues. In the proposed approach, the problem to be solved is the use of fast failover, which requires a large number of flow entries when a link failure occurs.

### B. Software Defined Networking

SDN provides a new approach to managing complex end-to-end connectivity and knowing the big picture of a network. A centralized network at the control layer allows management, configuration, security, and network resources to be optimized flexibly, dynamically, and automatically on SDN. It can be used for a variety of purposes, including control manipulation and network management, network virtualization, and providing a platform for building fast services [9].

As shown in Fig. 1, there is an SDN architecture consisting of an infrastructure layer related to the data plane that is in charge of forwarding. In the control layer, there is a component where the SDN controller is located. The application layer functions to make rules for the network. The control layer and the application layer are connected by the northbound API, while the infrastructure layer and the control layer are connected by the southbound API. In the southbound API, there is an SDN protocol, which is known as OpenFlow.

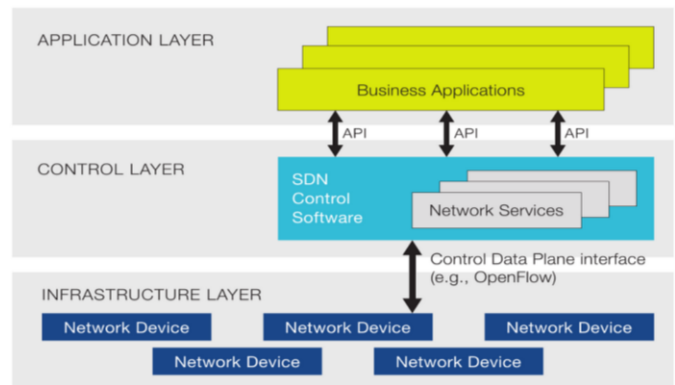


Fig. 1. SDN architecture

### C. OpenFlow

OpenFlow is a standard protocol used in software-defined networks. This protocol is used for communication between the control plane and the data plane. The SDN controller can manage a collection of switches to manage network traffic. The controller communicates with the OpenFlow switch and manages the switch via the OpenFlow protocol [10]. Fig. 2 shows the OpenFlow Controller and the OpenFlow Switch are the two most important components of OpenFlow. The OpenFlow Controller manages the performance of the Switch by controlling paths and flows. Then, the OpenFlow Switch is part of the data plane, which functions to process data such as forwarding packets.

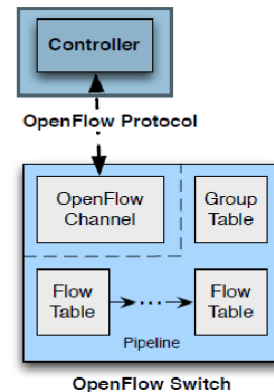


Fig. 2. Components of OpenFlow

In the OpenFlow protocol, each flow table on the switch has a flow entry, where each flow entry has a match field, a set of instructions to be applied to matching packets [11]. There are three main components in OpenFlow: the first is the table, which contains the flow table, meter table, and group table. The second is a secure channel that contains an SSL channel that is between the switch and the SDN controller. The third component is the OF protocol, which is used to control and manage switches [12].

### D. Failure Recovery

Failure recovery is a network process that allows packets or flows that have experienced link failure to be recovered and forwarded to their destination. In the failure recovery process, there are two mechanisms: reactive mechanisms and proactive

mechanisms [13]. In a reactive mechanism, there is no backup path configured in the forwarding plane, so the controller immediately computes an alternative path after receiving a link failure notification message from the switch. Whereas in a proactive mechanism, there are two separate paths (the primary path and the backup path) that are configured by the controller in the forwarding plane before link failure occurs on the network. The fast failover feature is used to implement a backup path on a proactive mechanism. Fast failover is the ability of a flow table to create a group table that provides various ways of forwarding (primary and backup links) [14]. With this capability, fast failover can redirect disrupted flows to a backup link that has been configured in the flow table. Fault recovery is performed directly by the OpenFlow switch without involving the controller. In addition, failure recovery can be combined with backup path calculations, which are proactively installed by the controller.

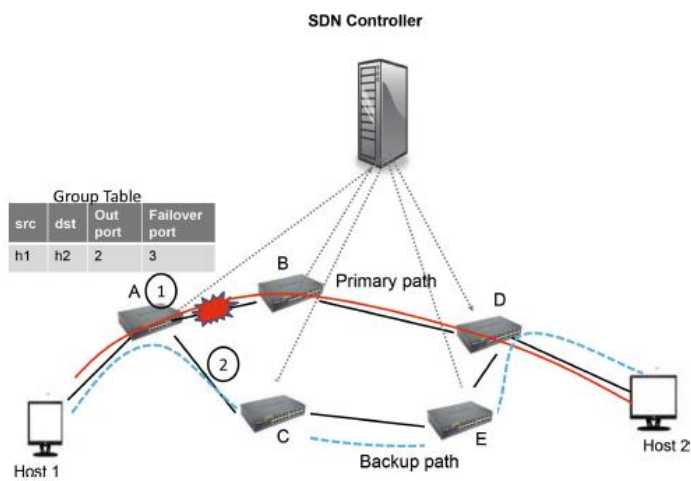


Fig. 3. Failure recovery mechanisms

Fig. 3 is an example of how the failure recovery mechanism works. There are 5 switches connected to 1 controller, and there are 2 hosts. In this topology, there is a primary path in switch A-B-D and a backup path in switch A-C-E-D. When there is a link failure on link A-B, packets from host 1 will be diverted to the backup path that has been configured in the flow table to switch A-C-E-D, so that packets can be sent to host 2.

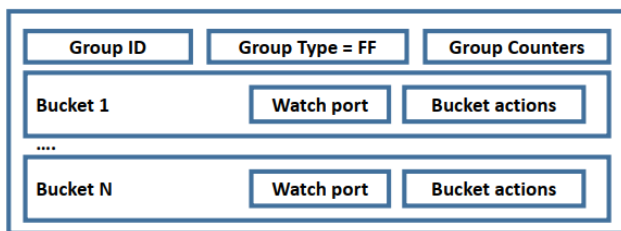


Fig. 4. Components of fast failover group

Fig. 4 shows the fast failover group component that is run by the SDN controller. Fast failover is a feature that reconfigures the link in the event of a failure. This feature utilizes OpenFlow 1.3 to run a group table that contains watch ports and action buckets that can monitor and act as long as the

port status changes [15]. The SDN controller configures the switch with a flow table that can help the network recover when a link failure occurs. The flow table contains fast failover group rules that implement a path-switching mechanism if a link is down.

### E. VLAN ID

VLAN-ID is one of the newest features introduced to OpenFlow in version 1.3. The VLAN mechanism can logically divide networks that are grouped based on VLAN ID. With this capability, VLANs can limit broadcast traffic on the network because they can only send packets to hosts that have the same VLAN ID. IEEE 802.1Q is the standard definition of VLANs. A VLAN tag contains 12 bits in the ethernet frame, so there can be up to 4,096 VLANs on a LAN. Implementation of VLAN tagging on the Ethernet protocol can create different broadcast sub-domains on the same LAN by including a VLAN number or tag for each subnet interface on the same switch [16].

In the OpenFlow protocol, in the flow table, there are flow entries that determine how a flow is processed and forwarded. Inside the flow entry, there are matching fields, actions, and counters. The matching field is used to match incoming packets. The action contains a set of instructions that are used to forward packets in various ways, for example, forwarding to a group table, one of which is the fast failover group. Then the counter is used to collect statistics on a particular flow, for example, the number of packets that have been received, the number of bytes, and the duration of the flow.

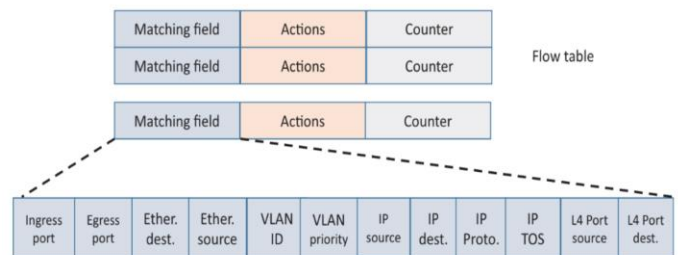


Fig. 5. Fields in the OpenFlow protocol

As shown in Fig. 5, there are 12 match fields, which are collectively referred to as the "basic twelve-tuple of match fields". Flow entries are processed sequentially, and when a match is found, the matching process against the flow table will be stopped [17]. In addition, the flow table is also equipped with a frame/byte counter that provides an indication of flow statistics on all ports so that the controller knows all the conditions that occur in the network [18]. Several actions can be performed by the OpenFlow protocol, such as sending packets to several ports, adding, removing, or modifying a VLAN tag, deleting packets, or sending packets to the controller.

### III. PROPOSED APPROACH

In OpenFlow, a specific flow can be defined as a collection of matching fields. Therefore, VLAN ID can be used as a flow ID, which can be forwarded based on flow entry. The use of VLAN ID can reduce interference with route flow and thus

reduce switch memory consumption. The failure recovery mechanism proposed in this research uses the VLAN ID feature on OpenFlow to collect failed flows. Each switch and a link is associated with a VLAN ID that can redirect flow to a backup path configured by fast failover on the switch.

Based on the topology that will be used in this research, the failure recovery mechanism will be implemented using OpenFlow 1.3. This mechanism provides a primary path for

forwarding and a backup path for diverting packets to an alternative path when a link failure occurs in the primary path. Based on Fig. 6, shows the failure recovery mechanism in the topology. In this topology, there is a primary path in S1-S2-S3-S4, while the backup path is in S2-S5-S6-S7-S4. Then, with the backup path configuration that has been made in S2, traffic can be diverted from port 3 to S5-S6-S7-S4 without making a round trip to the controller, so that packets can be sent to the destination.

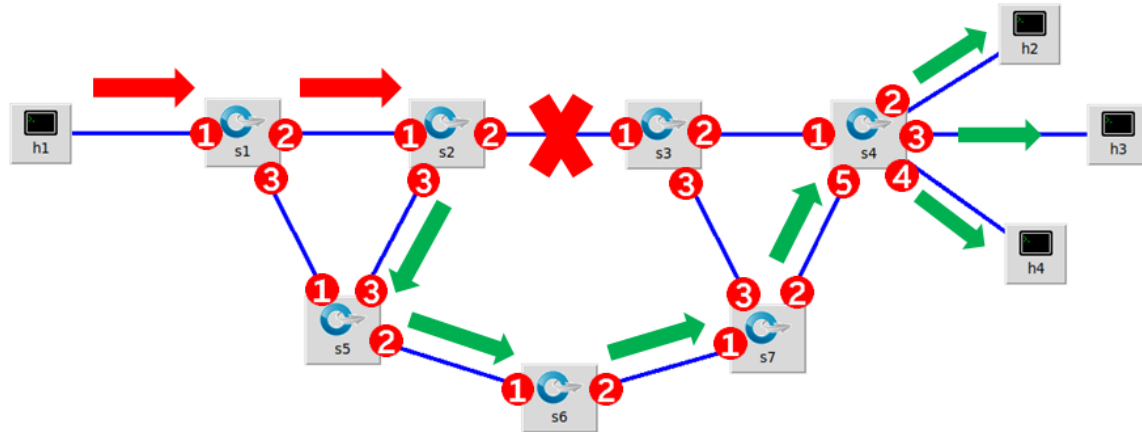


Fig. 6. Experimental topology

In the network topology, there is a link failure in S2-S3, so packets cannot pass through the link. Furthermore, there is H1, which will send packets to H2, H3, and H4, so there are three traffic flows in the topology. When the S2-S3 link fails, the controller will update the network topology by removing the failed link. In flow table S2, traffic is forwarded to group table 2 with the fast failover type, which is sent to output port 3.

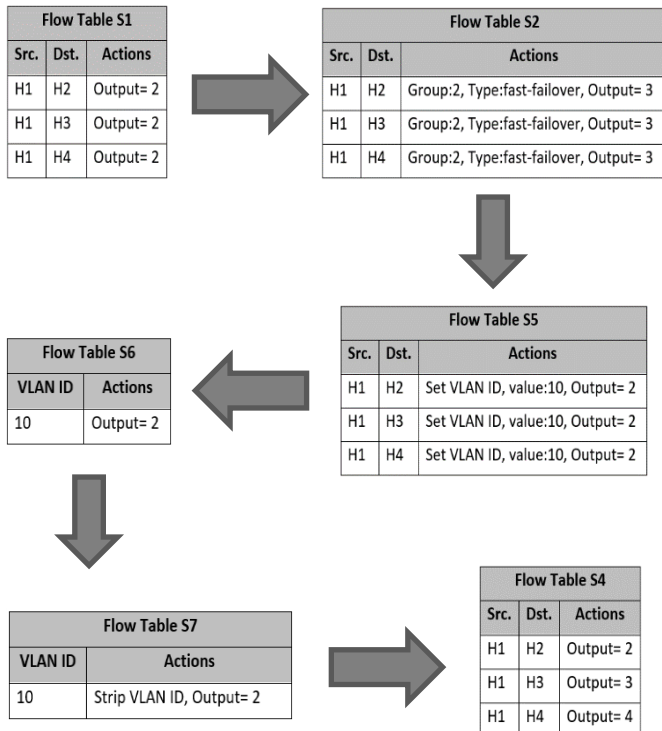


Fig. 7. Labelling process of VLAN ID

In Fig. 7, when the packet arrives at S5, configure the VLAN ID in the access port by accepting all packets that do not have a VLAN header. Then add a VLAN ID tag with a value of 10 for each incoming packet in S5. Then, the value of the VLAN ID is used as a match field in the switch connected to S5 and S6 via port 2. When it arrives at S6, packets will become one flow with a match field VLAN ID of 10. In S6, there is a packet with a VLAN ID as a match field with a value of 10 that has been configured on the previous switch. When a packet with a VLAN ID matches the flow match field, the packet can be forwarded based on the action specified in the flow table. Then, when the flow arrives at S7, there is an action with the Strip VLAN ID that functions to delete the value from the VLAN ID. The process is in the output access port, so the VLAN header has been deleted when it goes to the output port. Furthermore, when the flow arrives at S4, it will be returned to three traffic flows. Thus, in flow table S4, the three traffic rules in S4 without requiring changes to the flow table S4.

#### IV. PERFORMANCE EVALUATION

##### A. Testbed Configuration

The topology shown in Fig. 6 was implemented on Mininet as the network emulator and select Ryu as the controller. Mininet supports different types of switches. In this case, we used OpenVSwitch to support the fast failover group and VLAN ID features in OpenFlow 1.3. Furthermore, because our experiment was carried out in a controlled environment, we used OpenVSwitch to install the flows directly in each switch of the network topology using the script-line program *ovs-ofctl*. To collect statistics and monitor the behavior of TCP traffic generated by the IPERF application. The main characteristics

of the laptop on which the tests were conducted are as follows: Processor: Intel Core i7-8550U, CPU running at 1.99 GHz; RAM: 16 GB; operating system: Linux Ubuntu 20.04 LTS 64-bit on a VMware workstation.

### B. Recovery Time

The recovery time evaluation is carried out to find out how long it takes for packets to be diverted to the backup path when a link fails. The analysis compares the recovery time in failure recovery with VLAN ID and fast failover.

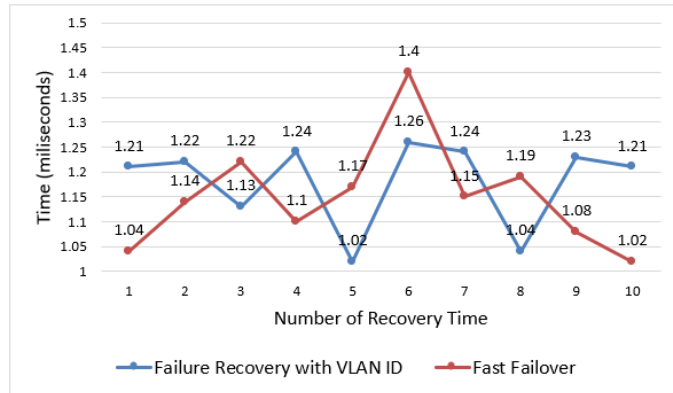


Fig. 8. Failure recovery time

In Fig. 8, there is a graph of the recovery time test. Based on the test results, the minimum time required for failure recovery with a VLAN ID is 1.02 ms, while the maximum time is 1.26ms. As a result, the time required to perform recovery in this mechanism is 1.02-1.26ms. Whereas for fast failover, the minimum time is 1.02ms, and has the maximum time of 1.4ms. As a result, when a link fails, recovery takes 1.02-1.4 ms on fast failover. Based on the results obtained in this test, the recovery time required in the failure recovery mechanism with VLAN ID is smaller than that required in the fast failover mechanism.

### C. Traffic Packets

The Traffic packet evaluation is used to find out how much packet traffic is transmitted to the destination. The analysis of this evaluation is used to determine the performance of the VLAN ID as a matching field in sending packets to the destination.

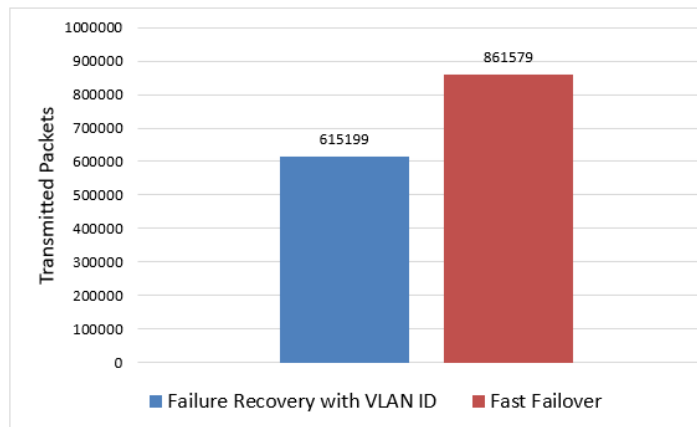


Fig. 9. Total number of traffic packets

As shown in Fig. 9, there is a traffic packet evaluation. Based on the two mechanisms tested, the failure recovery mechanism with VLAN ID resulted in a total of 615,199 traffic packets. Meanwhile, in failure recovery, the number of packets generated in this test was 861,579 packets. According to the results obtained from the test, failure recovery with VLAN ID produces less packet traffic than the fast failover mechanism.

### D. Packet Loss

Packet loss evaluation is carried out to find out how many packets are lost when sending packets from host 1 to host 2 when a link failure occurs. The duration of each test to be carried out is 10 seconds, and the test is carried out five times. The test will be carried out with a different total number of streams.

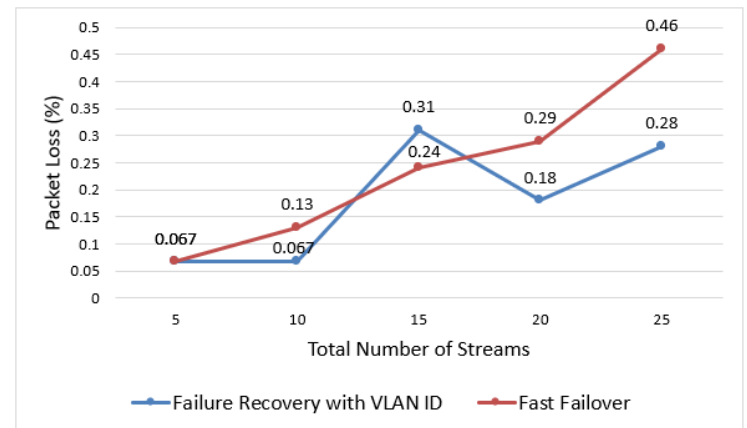


Fig. 10. Packet loss rate

In Fig. 10, there is a graph of the results of the packet loss evaluation that has been done. In the first test using 5 streams, the failure recovery mechanism with VLAN ID has a packet loss value of 0.067%, and this value increases when the last test uses 25 streams at 0.28%. Whereas in fast failover, in the first test, it has the same packet loss value of 0.067%. But in the last test using 25 streams, the packet loss value was 0.46%. Based on the results obtained in this test, the smallest packet loss value for the two mechanisms is 0.067%. Whereas in the last test, the failure recovery mechanism with VLAN ID had a smaller packet loss value compared to fast failover.

## V. DISCUSSION

In the recovery time test results, the failure recovery mechanism with VLAN ID shows results of 1.02-1.26ms with an average yield of 1.18ms to perform recovery after a link failure occurs. In research conducted [19], the time needed to detect and recover a single link failure is at least around 10-20 ms. Then, research [20], states that the need to detect and perform recovery on operator-scale networks must be carried out in 50ms time intervals. The results of the packet loss test that has been carried out show that the average value of packet loss in the failure recovery mechanism with a VLAN ID is 0.18%, with the highest packet loss being 0.28%. The results of the packet loss test are still considered good, based on research conducted [21] which states that packet loss with a ratio of 5-10% can affect network quality. Whereas in audio and video

stream scenarios, the range of acceptable packet loss is between 1 and 2.5%.

## VI. CONCLUSION

Based on the results of the analysis of this research it can be concluded that compared to the fast failover method used in SDN, it mainly has three advantages: First, in the recovery time test results, the failure recovery mechanism with VLAN ID shows results of 1.02-1.26ms to perform recovery after a link failure occurs. Whereas in the fast failover mechanism, the time needed to perform recovery is 1.02-1.4ms. Second, it shows that the use of VLAN ID in failure recovery is proven to be able to reduce the amount of traffic packet when a link failure occurs. Third, the results of the packet loss evaluation that has been carried out show that the average value of packet loss in the failure recovery mechanism with a VLAN ID is 0.18%, with the highest packet loss is 0.28%. Based on the evaluation results, our proposed approach has better results than the fast failover method. However, the major drawback of our proposed approach is that the mechanism is less dynamic because we implement fast failover groups and VLAN IDs directly in the switch. Perhaps we can present a solution to the problem and provide direction for our future work.

## REFERENCES

- [1] D. Kreutz, F. M. V. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proceedings of the IEEE*, vol. 103, no. 1, pp. 14–76, Jan. 2015, doi: 10.1109/JPROC.2014.2371999.
- [2] V. Muthumanikandan and C. Valliyammai, "A survey on link failures in software defined networks," *ICoAC 2015 - 7th International Conference on Advanced Computing*, Sep. 2016, doi: 10.1109/ICOAC.2015.7562808.
- [3] R. Kanagavelu and Y. Zhu, "A pro-active and adaptive mechanism for fast failure recovery in SDN data centers," *Advances in Intelligent Systems and Computing*, vol. 886, pp. 239–257, 2019, doi: 10.1007/978-3-030-03402-3\_17/COVER.
- [4] V. Muthumanikandan and C. Valliyammai, "Link Failure Recovery Using Shortest Path Fast Rerouting Technique in SDN," *Wireless Personal Communications 2017 97:2*, vol. 97, no. 2, pp. 2475–2495, Jun. 2017, doi: 10.1007/S11277-017-4618-0.
- [5] S. Petale and J. Thangaraj, "Link Failure Recovery Mechanism in Software Defined Networks," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 7, pp. 1285–1292, Jul. 2020, doi: 10.1109/JSAC.2020.2986668.
- [6] L. Huang, Q. Shen, and W. Shao, "A source routing based link protection method for link failure in SDN," *2016 2nd IEEE International Conference on Computer and Communications, ICC 2016 - Proceedings*, pp. 2588–2594, May 2017, doi: 10.1109/COMPComm.2016.7925166.
- [7] M. H. Wang, S. Y. Wu, L. H. Yen, and C. C. Tseng, "PathMon: Path-specific traffic monitoring in OpenFlow-enabled networks," *International Conference on Ubiquitous and Future Networks, ICUFN*, vol. 2016-August, pp. 775–780, Aug. 2016, doi: 10.1109/ICUFN.2016.7537143.
- [8] S. Chakraborty and C. Chen, "A low-latency multipath routing without elephant flow detection for data centers," *IEEE International Conference on High Performance Switching and Routing, HPSR*, vol. 2016-July, pp. 49–54, Jul. 2016, doi: 10.1109/HPSR.2016.7525638.
- [9] C. Decusatis et al., "Dynamic, software-defined service provider network infrastructure and cloud drivers for SDN adoption," *2013 IEEE International Conference on Communications Workshops, ICC 2013*, pp. 235–239, 2013, doi: 10.1109/ICCWork.2013.6649235.
- [10] F. Hu, Q. Hao, and K. Bao, "A survey on software-defined network and OpenFlow: From concept to implementation," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 4, pp. 2181–2206, Apr. 2014, doi: 10.1109/COMST.2014.2326417.
- [11] A. Mukherjee, R. A. Saeed, S. Dutta, and M. K. Naskar, "Fault tracking framework for software-defined networking (SDN)," *Resource Allocation in Next-Generation Broadband Wireless Access Networks*, pp. 247–272, Feb. 2017, doi: 10.4018/978-1-5225-2023-8.CH011.
- [12] S. Jamali, A. Badirzadeh, and M. S. Siapoush, "On the use of the genetic programming for balanced load distribution in software-defined networks," *Digital Communications and Networks*, vol. 5, no. 4, pp. 288–296, Nov. 2019, doi: 10.1016/J.DCAN.2019.10.002.
- [13] R. Ahmed, E. Alfaki, and M. Nawari, "Fast failure detection and recovery mechanism for dynamic networks using software-defined networking," *Proceedings of 2016 Conference of Basic Sciences and Engineering Studies, SGCAC 2016*, pp. 167–170, Apr. 2016, doi: 10.1109/SGCAC.2016.7458023.
- [14] E. Molina, E. Jacob, J. Matias, N. Moreira, and A. Astarloa, "Using Software Defined Networking to manage and control IEC 61850-based systems," *Computers & Electrical Engineering*, vol. 43, pp. 142–154, Apr. 2015, doi: 10.1016/J.COMPELEENG.2014.10.016.
- [15] K. Halba, C. Mahmoudi, and E. Griffor, "Robust safety for autonomous vehicles through reconfigurable networking," *Electronic Proceedings in Theoretical Computer Science, EPTCS*, vol. 269, pp. 48–58, Apr. 2018, doi: 10.4204/EPTCS.269.5.
- [16] M. B. Lehocine and M. Batouche, "Flexibility of managing VLAN filtering and segmentation in SDN networks," *2017 International Symposium on Networks, Computers and Communications, ISNCC 2017*, Oct. 2017, doi: 10.1109/ISNCC.2017.8071999.
- [17] P. Göransson, C. Black, and T. Culver, "The OpenFlow Specification," *Software Defined Networks*, pp. 89–136, Jan. 2017, doi: 10.1016/B978-0-12-804555-8.00005-3.
- [18] G. Pujolle, "Software Networks: Virtualization, SDN, 5G, and Security," Wiley eBooks, 2020. <https://ieeexplore.ieee.org/book/9116614> (accessed Oct. 03, 2022).
- [19] S. Sharma, D. Staessens, D. Colle, M. Pickavet, and P. Demeester, "Enabling fast failure recovery in OpenFlow networks," pp. 164–171, Dec. 2011, doi: 10.1109/DRCN.2011.6076899.
- [20] D. Staessens, S. Sharma, D. Colle, M. Pickavet, and P. Demeester, "Software defined networking: Meeting carrier grade requirements," *IEEE Workshop on Local and Metropolitan Area Networks*, 2011, doi: 10.1109/LANMAN.2011.6076935.
- [21] M. Pundir and J. K. Sandhu, "A Systematic Review of Quality of Service in Wireless Sensor Networks using Machine Learning: Recent Trend and Future Vision," *Journal of Network and Computer Applications*, vol. 188, p. 103084, Aug. 2021, doi: 10.1016/J.JNCA.2021.103084.

# Analysis of the Artificial Neural Network Approach in the Extreme Learning Machine Method for Mining Sales Forecasting Development

Hendra Kurniawan<sup>1</sup>, Joko Triloka\*<sup>2</sup>, Yunus Ardhan<sup>3</sup>

Information System-Faculty of Computer Science, Institut Informatika dan Bisnis Darmajaya, Lampung, Indonesia<sup>1,3</sup>  
Informatics Engineering-Faculty of Computer Science, Institut Informatika dan Bisnis Darmajaya, Lampung, Indonesia<sup>2</sup>

**Abstract**—Forecasting is an accurate indicator to support management decisions. This study aimed to mining sales forecasting on Indonesia's consumer goods companies with business warehouses engaged in the dynamic movement of large data using the Artificial Neural Network method. The sales forecasting used traditional method by inputting data and improvising simple patterns by collecting historical sales and remaining stock. Furthermore, several data variables in business warehouses were employed for sales forecasting. The study also used qualitative method to investigate the quality of data that cannot be measured quantitatively. The results showed with Mean Square Error score of 0.02716 in forecasting sales. The average accuracy generated by the Extreme Learning Machine after nine data tests is 111%. The result shows an opportunity for the company to further analyze the sales profit growth potential. The predicted value generated by Extreme Learning Machine for the last three months reaches 132%. The company's improved decision-making enlarge potential production line demonstrates the usefulness of this study.

**Keywords**—Artificial neural network; business warehouse; extreme learning machine; mining sales forecasting

## I. INTRODUCTION

Sales forecasting is always at the forefront of decision-making and planning [1] and also for business [2]. Forecasting gives an organization the right plan to deal with future demands but does not guarantee the success of a strategy. However, its failure results of wrong decisions in marketing activities caused by the allocation of other resources which constructed on imprecise and uncertain assumptions, hence resulting to wrong decisions [3]. Sales forecasting is an important prerequisite in many aspects of sales chain management. Therefore, further optimization efforts are needed to create sales forecasts that support decision-making by the organization.

This study aimed to explore the potential of scientific forecasting and prove the optimization of existing theories. The forecasting concept uses the *Artificial Neural Network* (ANN) method, representing the human brain's performance [4]. The human brain always experiences a period of learning in the interconnected neurons. Information received by neurons is sent from one neuron layer to another [5].

This study also explained the *Extreme Learning Machine* (ELM) which is widely used in batch, sequential, and incremental learning. The method is used because of its

efficient convergence and learning speed, good generalization ability, and ease of implementation [6]. Furthermore, the study built a suitable model to estimate the data generation process underlying the series. It also estimated the desired number of future observations through this model [6].

Several data variables in the business warehouse were adopted for sales forecasting using the ANN ELM method. Various studies have revealed that the ELM method has an advantage in learning speed which described in the previous section. It helps determine the effectiveness of forecasting, as well as deficiencies and necessary improvements.

## II. RELATED WORK

Prianda and Widodo (2021) used ELM to project foreign tourist visits and obtained a forecast MAPE value of 7.62% [7]. Similarly, Sharma et al. [8] employed an intelligence model to predict sales using sentimental product analysis. The study made the model more productive in obtaining accurate results for each item in the product life cycle. Meanwhile, [9] applied ELM integrated with LSTM to Bitcoin price forecasting. Their study matches the different machine learning algorithms to corresponding multiscale components and constructs the ensemble prediction models based on machine learning and multiscale analysis. The results showed that the ensemble models can achieve a prediction accuracy of 95.12% with enhanced performance than the benchmark models [9]. Moreover, Cholid and Aly [10] forecasted spiral and leaf spring products for four-wheeled vehicles using the Artificial Neural Network Backpropagation method. The study employed a learning rate weight value of 0.1 of four hidden layers with error of 0.01. Several studies use various methods to predict product sales [11]–[13], compare forecasting techniques for financial prediction [14], supply chain [15]–[17], and manufacturing processes [18].

## III. METHOD

This study aimed to predict sales using the ELM method, which applies qualitative approaches. It also intended to investigate, discover and explain the peculiarities of the data influence that cannot be measured or described quantitatively. Fig. 1 presents the proposed methods used in this study.



The proposed method in Fig. 1 consists of five phases. Firstly, problem formulation was formed through the question regarding implementation of forecasting with other fast and accurate methods. After finding the problem formulation, a literature study was performed by collecting sources from previous journals and articles and quarrying up information from company forecasting results.

Furthermore, the ELM method accomplished with the first stage is initializing the input weight and bias to prevent layer activation outputs from detonating or fading gradients during the normalization. Second stages are use of range between 0-1 for normalizing to ensure quality of data during training and testing phase. Training and testing stage then taking into account respectively to train model, discover and learn patterns. Testing data use to evaluate the functioning and improvement of algorithms training and optimize it for better results. The last stage of ELM is denormalizing previous prediction results to compare the output in order to evaluate the model. The final step was to exam the *Mean Square Error* (MSE) using python pseudocode to determine the error rate of the prediction results.

For training and testing of the proposed model, the company's secondary data have been used. From the proposed model, it can be seen that the prediction result process quickly for both training and testing sets. Hence, the proposed model will provide more suitable for the task at forecasting sales compare to other techniques that described in the previous section.

A. Data Type and Source

This study used secondary data from the dataset of a retail company. The data were collected from daily sales based on the company's existing business warehouse system.

B. Data Used

The study used sales data set digitized through the company's SAP Business Warehouse system. The historical data set had 43 rows and 6 parameter features. Furthermore, a time series analysis was conducted on the previous month of sales data. This study also analyzed time, brand, and stock factors affecting sales levels through forecasting methods. The dataset design has been given in Table I.

TABLE I. DATASET DESIGN

1st Data (Time Series)	Feature Stock	Feature 1 Sales	Feature 2 Sales 2	Feature 3 Sales 3	Feature 4 Sales 4
Nov 2017	Vol	Vol	Vol	Vol	Vol
Time Series n	Vol	Vol	Vol	Vol	Vol

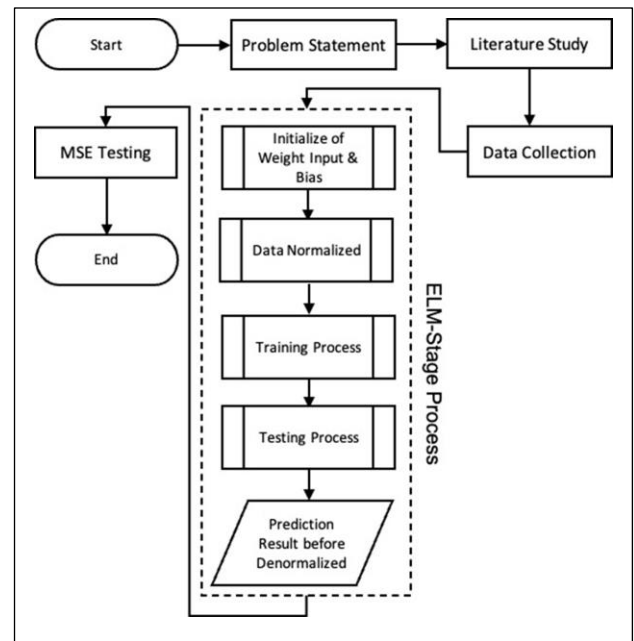


Fig. 1. Proposed forecasting sales through ELM stages

C. Schematic of Research Based on Method

The ELM algorithm was used by entering historical sales data, the number of features and hidden neurons, as well as the percentage of training and testing data as initial input. Fig. 2 shows the stages outlined as follows:

- a) Input weight and bias were processed randomly with values ranging between -1 to 1 based on the number of neurons then. The next step was transposing the matrix and normalizing the data using the range 0 - 1 to calculate the hidden layer output.
- b) The matrix result was multiplied by the transposition before calculating the Moore-Penrose Generalized Inverse matrix using the Moore-Penrose Generalized Inverse matrix equation. The result was multiplied by the transposition of the activated hidden layer output matrix.
- c) The study calculated the output results of the training process to obtain the output weight used in the testing process.
- d) The testing process employed the input weight and bias obtained from the training process. The hidden layer output was then calculated using the activation function.
- e) The output weight value obtained in the training process was used in the testing process to calculate the output layer as the prediction result.
- f) Before the denormalization process, the study calculated the error value on all output layers not denormalized with actual data. This error value is the prediction results obtained. Additionally, the error value was tested using Mean Square Error (MSE).
- g) The last step was denormalization to generate a previously denormalized value back to the original value. Fig. 2 is a flowchart of the problem-solving process with ELM.

D. Input Data

This study employed sales data detailing the monthly sales for November 2017-May 2021. The ELM method processes data by determining several criteria parameters to achieve a small error rate and optimal accuracy. Table II describes several parameters used in processing the dataset, including the input and output layers, hidden neurons, and activation function.

E. ELM Development Method Forecasting

The initial step was to find the optimal value of several input parameters tested. This step ensures that the process in the ELM method produces good predictions during training. After testing the MSE on hidden neurons, the dataset was normalized into the range of 0 – 1. This was followed by data training, testing, denormalization, and sales predictions in the following month.

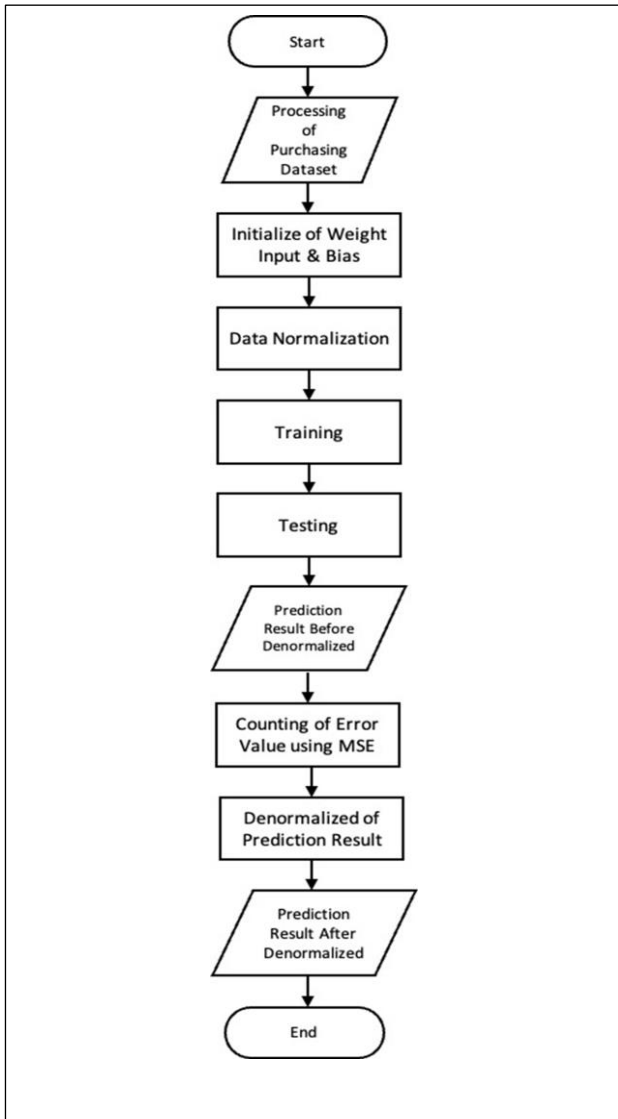


Fig. 2. ELM process diagram

TABLE II. ELM METHOD REQUIREMENT PARAMETERS

Parameter	Total	Description
Input Layer	6	Streak months, Data Stock (Label) & 4 Brand Historical Sales
Output Layer	1 Neuron	Forecasting Results
Hidden Layer	10 Testing	11 - 30 Neuron
Activation Function	1	Sigmoid Biner

IV. IMPLEMENTATION AND RESULT

A. Hidden Neuron Total Network Testing

The number of hidden neurons was tested to determine its effect on the accuracy value in implementing the ELM algorithm. The number of hidden neurons used in this test includes 11, 12, 13, 14, 15, 17, 24, 26, 27, and 30.

The tests were performed ten times, and the number of hidden neuron networks was processed into python pseudocode. Furthermore, the test was carried out repeatedly by changing the number of neurons, as shown in Table III.

TABLE III. HIDDEN NEURON TEST RESULTS

No	Total Hidden	MSE	Time (second)
1	11	0.06755	0.0515
2	12	0.02716	0.0525
3	13	0.04812	0.0528
4	14	0.08318	0.0545
5	15	0.10937	0.0526
6	17	0.12538	0.0532
7	24	0.13625	0.0550
8	26	0.21435	0.0548
9	27	0.18092	0.0545
10	30	0.24473	0.0547

Fig. 3 shows that changes occur when there is a change in the total of hidden layers. Therefore, the optimal network value was obtained in the second test with an MSE of 0.02716 and a time of 0.00525 seconds.

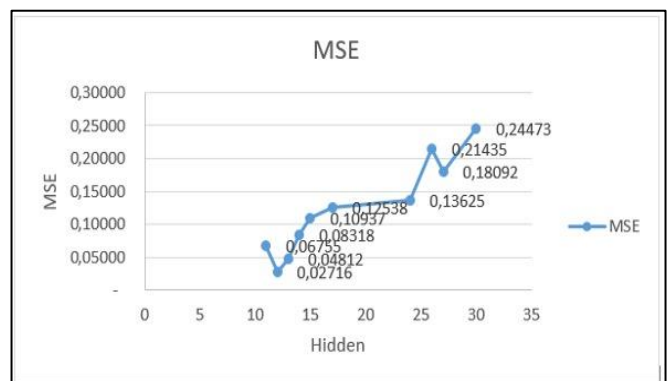


Fig. 3. Optimal hidden network graph plot

### B. Preprocessing

Data processing was performed after creating the hidden neuron network. The data collected were then normalized using Min-Max normalization or transformed into a range of 0-1 to obtain a value for each auxiliary variable.

The input process used monthly time series, stock, and sales variables by brand. The data were divided for testing and training via Python's `train_test_split` function. Furthermore, the study compared the performance of the models used in forecasting sales data. The training and testing data distribution was 80% and 20%, respectively. Tables IV and V show that the output data were broken down into two parts and formed a range of numbers 0-1 equated with numbers 0-1.

TABLE IV. DISTRIBUTION OF 80% OF TRAINING DATA AFTER NORMALIZATION

Feature 1	Feature 2	Feature3	Feature4	Feature5	Feature6
0.3552	0.3497	0.3589	0.3498	0.7480	0.7270
0.5395	0.3603	0.4630	0.3677	0.7573	0.7980
0.4781	0.3504	0.4128	0.3667	0.6723	0.7193
0.5268	0.3705	0.5047	0.3696	0.6141	0.6674
0.5049	0.3655	0.4704	0.3642	0.7085	0.5706
0.4635	0.3553	0.3978	0.3701	0.9262	0.6889
0.5078	0.3709	0.4193	0.3669	0.5677	0.6189
0.4899	0.3538	0.4004	0.3754	0.6889	0.6091
0.4777	0.3507	0.3950	0.3784	0.6787	0.6886
0.4399	0.3614	0.4304	0.3541	0.5897	0.6004
0.4511	0.3546	0.3994	0.3606	0.7527	0.6141
0.4414	0.3625	0.4619	0.3610	0.6547	0.5677
0.4758	0.3505	0.4119	0.3642	0.7157	0.6713
0.4537	0.3506	0.3798	0.3679	0.6708	0.6994
0.5057	0.3518	0.3929	0.3873	0.7114	0.6157
0.4960	0.3660	0.4483	0.3751	0.6091	0.6229
0.4642	0.3492	0.3900	0.3710	0.7454	0.6838
0.4565	0.3532	0.4274	0.3662	0.6004	0.6152
0.4644	0.3634	0.4306	0.3645	0.6813	0.6149
0.4722	0.3555	0.4084	0.3745	0.6229	0.6813
0.4781	0.3703	0.4697	0.3596	0.6935	0.7085
0.4448	0.3570	0.4504	0.3636	0.6432	0.5754
0.4375	0.3494	0.3671	0.3833	0.7270	0.7254
0.4868	0.3598	0.4206	0.3821	0.6713	0.9262
0.4590	0.3533	0.4360	0.3651	0.6152	0.6547
0.4251	0.3489	0.3942	0.3684	0.7193	0.7157
0.4282	0.3561	0.4971	0.3613	0.5754	0.6646

Feature 1	Feature 2	Feature3	Feature4	Feature5	Feature6
0.4595	0.3654	0.4490	0.3643	0.7980	0.7527
0.4700	0.3495	0.3972	0.3666	0.6838	0.6723
0.4789	0.3494	0.4008	0.3781	0.7254	0.7114
0.4622	0.3492	0.3834	0.3693	0.7294	0.7454
0.5021	0.3523	0.4390	0.3694	0.6457	0.7310
0.4964	0.3505	0.3859	0.3760	0.6886	0.7294
0.4651	0.3507	0.3809	0.3786	0.6157	0.6708

TABLE V. DISTRIBUTION OF TESTING DATA 20% AFTER NORMALIZATION

Feature 1	Feature2	Feature3	Feature4	Feature5	Feature6
0.48451	0.35353	0.42715	0.34968	0.69842	0.72659
0.43501	0.35957	0.42615	0.34449	0.64512	0.36723
0.45851	0.34328	0.40150	0.36011	0.69533	0.67470
0.46529	0.36314	0.48058	0.46193	0.76199	0.76036
0.47845	0.34694	0.45601	0.35457	0.76622	0.79335
0.49427	0.36567	0.50819	0.35478	0.59095	0.61250
0.48471	0.35135	0.46971	0.36826	0.64059	0.67189
0.52766	0.34101	0.40406	0.37653	0.73279	0.71063
0.49897	0.35415	0.41395	0.35304	0.36723	0.78375

### C. Preprocessing

The training process is conducted using 80% of the dataset. Table VI shows the data normalized and formed a training output table with Order 12 x 6. This implies 12 hidden layer networks and 6 input features.

TABLE VI. TRAINING ORDO 12 x 6

Feature1	Feature2	Feature 3	Feature 4	Feature 5	Feature 6
0.0558	0.8430	0.6263	0.4151	0.6627	0.0252
0.0896	0.6381	0.5356	0.7703	0.3389	0.7839
0.1758	0.3225	0.5412	0.7633	0.4020	0.5811
0.4652	0.5609	0.1406	0.4446	0.7704	0.6104
0.5294	0.6224	0.4036	0.7573	0.0588	0.1303
0.6689	0.3863	0.4126	0.8024	0.8661	0.5111
0.7164	0.6305	0.5334	0.3697	0.0297	0.4240
0.7621	0.1383	0.3566	0.2204	0.9654	0.0216
0.8333	0.2910	0.8613	0.3733	0.3512	0.2307
0.8409	0.7427	0.2408	0.0436	0.0281	0.1970
0.9815	0.1281	0.4837	0.3220	0.0628	0.5891
0.9848	0.5148	0.9862	0.9852	0.2135	0.8389

Randomized training data produce output weights needed during the testing process. Table VII show the weighted output.

TABLE VII. OUTPUT WEIGHT

Output Weight (Beta)
419.442941
414.794705
713.993594
-149.313835
-473.334959
30.800049
-260.807955
-146.883101
-248.036551
-35.957942
-295.455812
32.925058

*D. Testing*

The testing process aimed to measure the performance of the network model built during the training process. Although the steps used were similar to the training process, all weights were taken from the training results, implying no calculation of the  $\beta$  weight. The data used differed from the training process explained in the preprocessing stage. The accuracy level was calculated in the same way as in the training process. The output data testing is presented in Table VIII.

TABLE VIII. OUTPUT DATA TESTING

Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6
0.4845	0.3535	0.4272	0.3497	0.6984	0.7266
0.4350	0.3596	0.4261	0.3445	0.6451	0.3672
0.4585	0.3433	0.4015	0.3601	0.6953	0.6747
0.4653	0.3631	0.4806	0.4619	0.7620	0.7604
0.4785	0.3469	0.4560	0.3546	0.7662	0.7934
0.4943	0.3657	0.5082	0.3548	0.5909	0.6125
0.4847	0.3513	0.4697	0.3683	0.6406	0.6719
0.5277	0.3410	0.4041	0.3765	0.7328	0.7106
0.4990	0.3542	0.4139	0.3530	0.3672	0.7837

Table IX shows that the results generated from the training process had an average prediction of 111%.

TABLE IX. OUTPUT TARGET TESTING VS PREDICTION TESTING RESULTS

Data To-	20% target data (Y)	Prediction Results (y) Accuracy	
Data to-1	0.4057	0.5270	130%
Data to-2	0.2980	0.5998	201%
Data to-3	0.3145	0.5494	175%
Data to-4	0.7608	0.5992	79%
Data to-5	0.4882	0.5891	121%
Data to-6	0.6600	0.5091	77%
Data to-7	0.5859	0.5569	95%
Data to-8	0.5570	0.6899	124%
Data to-9	0.4179	0.3409	82%
Average	0.4987	0.5513	111%

*E. Denormalization*

In the denormalization stage, the predicted test data in the 0-1 output range were converted to the actual value in kilograms. This was done to ensure that the target data and predicted results were read on a wider scale than the original value.

The denormalized data in Table X results from nine data tested and predicted to compare each accuracy per line. The accuracy from the average of the final data line was 101%.

TABLE X. DENORMALIZATION OF TARGET TESTING

Data	20% Target Data (Y)	Prediction Results (y)	%
Data to-1	57.521	67.931	118%
Data to-2	47.425	74.235	157%
Data to-3	49.077	69.851	142%
Data to-4	90.227	74.183	82%
Data to-5	64.634	73.293	113%
Data to-6	79.734	66.413	83%
Data to-7	73.014	70.497	97%
Data to-8	70.499	82.638	117%
Data to-9	58.587	51.616	88%
Average	67.367	68.250	101%
Data to-7 - 9			

*F. Next Month Prediction Results*

The predictions made via pseudocode python showed that the value on the 44th data or the following month from the processed dataset is 70.524 kg, as shown in Fig. 4. In Fig. 4, the red line on the y-axis explains that sales occur in the 44th data, a decrease from the previous month.

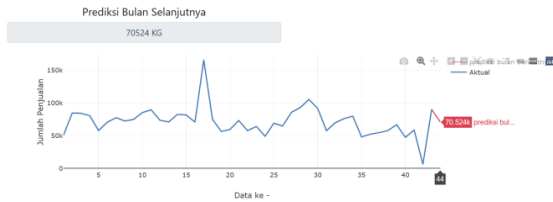


Fig. 4. The plot of the next month's prediction results

### G. Forecasting Exist Sales

Retail companies have a forecast to predict sales within the next three months. This is known through an analysis of traditional time series forecasting. Table XI and Table XII shows forecast existing and actual sales respectively. Table XIII compares the prediction made and the actual from the total forecasting, showing a gap with an accuracy of 86%.

TABLE XI. FORECAST EXISTING

Brands	March	April	May	Total
Promina	8.996	751	15.558	8.435
SUN	36.524	3.577	33.461	24.520
Govit	1.439	122	1.962	1.174
Gowell	2.844	1.233	21.270	8.449
Total	49.802	5.682	72.251	42.579

TABLE XII. ACTUAL SALES

Brands	March	April	May	Total
Promina	9.697	904	22.218	10.940
SUN	36.771	3.708	54.311	31.597
Govit	1.108	64	1.348	840
Gowell	11.010	1.553	12.349	8.304
Total	58.587	6.229	90.227	51.681

TABLE XIII. COMPARISON OF EXIST VS ACTUAL FORECASTING

Brands	March	April	May	Average
Forecast	49.802	5.682	72.251	42.579
Actual	58.587	6.229	90.227	51.681
Accuracy	85%	91%	80%	82%

Where:

*Forecast:* The forecasting performed by Retail companies in units of volume

Actual: Sales in March, April, and May

Accuracy: *Forecast/Actual Sales*

The average data for three months of prediction in Table XIII confirm that the average accuracy is only 82%. A less-than-actual plot accuracy warns the company of insufficient stock available to supply future sales. This means that sales would not be maximized.

### H. Comparison of Exist vs ELM Development Methods

*ELM exist vs development method matrix:* From the implementation of sales forecasting proposed model, a comparison matrix could be made between the Exist versus ELM Method as given in Table XIV. Based on the forecast matrix, it can be seen that the average 3-month forecast value reached 68,250 kilograms by using the ELM method, while the traditional method reached 42,579 kilograms. The average value of 3 months is actual sales 51,681 kilogram has been used for comparison the ELM and traditional method. Based on results, it can be concluded that predictions by ELM provide more better accuracy to provide adequate stock for the next sale. It mean that sales be able to run optimally for company. Meanwhile, based on accuracy matrix the ELM method can achieve admirable accuracy values, this is due to the stock feature as basic calculation label are carried out to acquire the smallest error value as well as several tests also has been done. Some of these features are support in providing stimulation to achieve better accuracy.

### V. CONCLUSIONS

Based on the test results and discussion, the following conclusions were obtained:

- As measured by the Mean Square Error (MSE) in forecasting sales results, the error rate is 0.02716 with a time of 0.0525.
- The ten neuron network tests showed that more hidden neurons do not measure algorithm optimization. Therefore, using 12 of 30 hidden neurons produces fewer error values.
- The average accuracy value generated by the ELM forecasting method when testing nine data is 111%. This illustrates an opportunity for the company to further analyze the sales profit growth potential.
- The predicted value produced by ELM for the last three months reaches 132% compared to traditional methods, which only achieve 82%.

TABLE XIV. COMPARISON OF THE EXIST VS ELM METHOD

Matrix	Method	Average Unit (Kg)	Description
Forecast	Exist (Traditional)	42.579	The average 3-month forecast value conducted by the company using the traditional method is 42.579 kilograms. This value is obtained from Table XI.
	ELM	68.250	The average value of the 3-month forecast using the ELM method is 68.250 kilograms. This value is obtained from Table XIII.
	Actual Average 3 Months	51.681	The average value of 3 months compared to actual sales is 51.681 kilograms. This value is obtained from Table XIII.
Accuracy	Exist (Traditional)	82%	Comparison of Exist / Traditional Forecast Values compared to Actual. Value calculated using spread sheet gets a relatively good accuracy value but has not yet reached the optimal value. This is because the calculated forecast still applies manual estimation as a measuring basis by dividing average monthly sales and adding sales estimates according to forecast growth.
	ELM	132%	Comparison of ELM Forecast Value compared to Actual. In line with the proof method described in the previous discussion, the accuracy value of the ELM method could achieve a very good accuracy value. This is due to the featured stock, the reference label for calculations, and several tests are conducted to get the smallest error value and the best accuracy. Some of these features help in providing stimulation to achieve optimal accuracy.

Sales forecasting still requires extensive machine learning and statistics knowledge. Addition features of input layer such as number of workers, demographic trends, and behavioral indicators should be involved in future studies to improve the model.

ACKNOWLEDGMENT

The authors thank KemendikbudRistek DIKTI Indonesia who provides funding of this work. The authors thank Institut Informatika dan Bisnis Darmajaya, especially the Faculty of Computer Science, for all the support.

REFERENCES

- [1] F. Petropoulos et al., "Review forecasting: theory and practice," Int J Forecast, vol. 38, no. 3, pp. 705–871, 2022.
- [2] O'Trakoun, J. "Business forecasting during the pandemic". Bus Econ vol. 57, pp. 95–110, 2022.
- [3] A. K. K. P. Dewi, "Penerapan metode extreme learning machine pada peramalan jumlah penumpang penerbangan domestik di Kepulauan Riau," in Prosiding Statistika, pp. 589–596, 2021.
- [4] A.G. Bonab, "A comparative study of demand forecasting based on machine learning methods with time series approach," Journal of Applied Research on Industrial Engineering, vol. 9, no. 3, pp. 331–353, 2022.
- [5] M. S. R. Dastres, "Artificial neural network systems," International Journal of Imaging and Robotics (IJIR), vol. 21, pp. 13–25, 2021.
- [6] N. A., R. M. J., M. and A. Z. B.P. Ooi, "A study of extreme learning machine on small sample-sized classification problems," J. Physics.: Conference Series, vol. 2107, pp. 1–7, 2021.
- [7] B. Prianda and E. Widodo, "Perbandingan metode seasonal ARIMA dan extreme learning machine pada peramalan jumlah wisatawan mancanegara ke Bali," Barekeng: J. Math. & App, vol. 15, pp. 639–650, 2021.
- [8] N. G. J. R. and M. B. A. K. Sharma, "An Intelligent model for predicting the sales of a product," in 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 341–345, 2020
- [9] C. Luo, L. Pan, B. Chen and H. Xu. "Bitcoin price forecasting: an integrated approach using hybrid LSTM-ELM models". Mathematical Problems in Engineering, vol. 2022, 2022.
- [10] D. A. F. Cholid, "Implementation of product sales forecast using artificial neural network method," International Journal of Information System and Technology (IJISTECH), vol. 5, pp. 153–162, 2021.
- [11] A. U. P. Paduloh, "Analysis and comparing forecasting results using time series method to predict sales demand on Covid-19 pandemic era," Journal of Engineering and Management in Industrial Systems, vol. 10, pp. 37–49, 2022.
- [12] P. Smirnov & S. Vladimir. "Forecasting new product demand using machine learning". Journal of Physics: Conference Series. 1925. (1): 012033, 2021
- [13] D. H. and R. A. N. V. Wineka, "Sales Forecasting by iusing exponential smoothing method and trend method to optimize product sales in PT. Zamrud Bumi Indonesia during the Covid-19 pandemic," Int. J. Eng. Sci. Inf. Technology, vol. 1, pp. 59–64, 2021.
- [14] M. U. G. and A. M. O. O. B. Sezer, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," Appl. Soft Comput. J, vol. 90, pp. 2005–2019, 2020.
- [15] E. Beh., G. T. and R. G. M. Abolghasemi., "Demand forecasting in supply chain: the impact of demand volatility in the presence of promotion," Comput Ind Eng, vol. 142, 2020.
- [16] S. S. W. S. and M. M. N. E. Fradinata, "Compare the forecasting method of artificial neural network and support vector regression model to measure the bullwhip effect in supply chain," J. Mech. Eng. Sci, vol. 13, pp. 4816–4834, 2019.
- [17] M. Alnahhal, D. Ahrens, and B. Salah, "Dynamic lead-time forecasting using machine learning in a make-to-order supply chain," Applied Sciences, vol. 11, no. 21, p. 10105, Oct. 2021.
- [18] R. Kannan., H. A. Abdul Halim., K. Ramakrishnan, et al. Machine learning approach for predicting production delays: a quarry company case study. J Big Data 9, no. 94, 2022.

# Deca Convolutional Layer Neural Network (DCL-NN) Method for Categorizing Concrete Cracks in Heritage Building

Dinar Mutiara Kusumo Nugraheni\*<sup>1</sup>, Andi Kurniawan Nugroho\*<sup>2</sup>, Diah Intan Kusumo Dewi<sup>3</sup>, Beta Noranita<sup>4</sup>  
Informatic Department-Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia<sup>1,4</sup>  
Electronic Department-Faculty of Engineering, Semarang University, Semarang, Indonesia<sup>2</sup>  
Urban Regional Planning Department-Faculty of Engineering, Universitas Diponegoro, Semarang, Indonesia<sup>3</sup>

**Abstract**—It is critical to develop a method for detecting cracks in historic building concrete structures. This is due to the fact that it is a method of preserving historic building and protecting visitors from the collapse of a historic structure. The purpose of this research is to determine the best method for identifying cracks in the concrete surface of old buildings by using cracked images of old buildings. The various surface textures, crack irregularities, and background complexity that distinguish crack detection from other forms of image detection research present challenges in crack detection of old buildings. This study presents a framework for detecting concrete cracks in old buildings in Semarang's old town using a modified Convolutional Neural Network with a combination of several convolutional layers. This study employs ten convolutional layers (Deca Convolutional Layer Neural Network (DCL-NN)) to provide mapping features for images of concrete cracks in ancient buildings at preservation area. This study also compares commonly used machine learning models such as KNeighbors (n neighbors=3), Random Forest, Support Vector Machine (SVM), ExtraTrees (n estimators=10), and other CNN-pretained models such as VGG19, Xception, and MobileNet. Four performance indicators are used to validate each model's performance: accuracy, recall, precision, F1-score, Matthews Correlation Coefficient (MCC), and Cohen Kappa (CK). This study's data set is comprised of primary data obtained from cracked and normal images of several buildings in Semarang's old town. The accuracy of this study using DCL-NN is 98.87%, recall is 99.40%, precision is 98.33%, F1 is 98.86%, MCC is 97.74%, and CK is 98.86% for crack class. From this study, it was found that the ten convolution layers have higher classification performance compared to other comparison models such as machine learning and other CNN models and are more effective in detecting cracks in concrete structures.

**Keywords**—Cracks; concrete; Deca-CNN; features mapping; performance

## I. INTRODUCTION

It is critical to understand the shape of cracks in historic buildings in order to preserve historic areas. The Old Town area of Semarang, which is a UNESCO World Heritage Site, is one area that is vulnerable to building cracks [1]. The total number of buildings, which reached 274 units, demonstrates that the old town area was previously a residential area; now,

out of a total of 157 units with the status of occupied buildings (both for housing and offices, dominated by offices), 87 units are status as vacant buildings (both those that are still being maintained or damaged/abandoned), 28 units are leased (offices), and only 2 units are currently sold [2]. Historic ancient buildings are architectural creations that serve as a nation's cultural heritage and have very high artistic and historical values, so their long-term viability must be ensured. The ability of a sturdy building structure supports the building's strength, which causes the building to last a long time. Cracks in concrete structures are a common sign of faulty concrete. The presence of cracks affects the structural condition and increases the risk of unexpected damage and collapse of the building [3][4]. Therefore, crack detection must be done on a regular basis in order to maintain the concrete structures of historic ancient buildings.

## II. RELATED WORKS

Traditionally, professional images of cracks in a building's concrete structure are used. The use of an expert to inspect concrete structures for cracks is costly, time-consuming, and sometimes dangerous for direct inspection [5]. The Ultrasonic Pulse Velocity Test (UPVT) in the form of ultrasonic waves in the cracks where holes are made for the ultrasonic wave propagation area is then used as another method of measuring building cracks [6]. This creates the possibility for buildings to become more dangerous as a result of the influence of other building structures. However, as computer-aided design (CAD)-based image processing technology advances, many experts are turning to machine learning-based image processing for the automatic detection of cracks in concrete structures [7]. Many techniques for detecting building cracks have been proposed by researchers, including the use of thresholding methods [8], edge detection, and wavelet transform. Surface texture, crack irregularity, and background complexity distinguish crack detection from other images in research that leads to machine learning-based image processing solutions for automatic detection of cracks in concrete structures.

Deep learning-based models, especially multilayer neural networks, currently play an important role in feature learning [9]. Moreover, the availability of high-performance computers and the continuous improvement of good training methods on

\*Corresponding Author.

Universitas Diponegoro Research Publication International {RPI} Grant  
SK. Rector Universitas Diponegoro No: 215 /UN7.A/HK/VII/2022;  
No.SPK:569-108/UN7.D2/PP/VII/2022

available datasets are driving the rapid development of deep learning. Conversely, convolutional neural networks (CNNs) are feed-forward neural networks characterized by high-resolution image processing [10]. Some of these models are suitable for feature extraction in various applications, but their accuracy needs to be improved to detect cracks in concrete.

In this work, CNN-based transfer learning using pre-trained models to achieve efficient performance, reduce training time, overcome the drawbacks of large datasets, and yield significant results. A law has been proposed [11]. Some previous researchers have done some work to identify and classify cracks in buildings. Zhang et al. [12] proposed a 6-layer convolutional neural network (CNN) architecture for road crack detection and used 640,000, 160,000, and 200,000 images to train, validate, and validate the network. bottom. and tested. Kings. [13] proposed a CNN architecture with three convolutions and he two fully connected layers to detect cracks in asphalt. On the other hand, the author of this study says that for training he used 640,000 images and for testing he used 120,000 models. Fan et al. [14] proposed an efficient automatic road crack detection and measurement model based on an ensemble of CNN models. The authors of this study calculated the final failure probability by combining the probability values from each CNN model using the weighted overall average technique. Xu et al. [15] He trained a 28-layer end-to-end CNN model to detect cracks in concrete bridges. To obtain multiscale contextual information, the authors of this study used the concept of combining Atlas Spatial Pyramids (ASPP) and depth convolution to reduce the number of parameters in the network. This study describes a framework for detecting concrete cracks in old buildings in Semarang Old Town using a modified convolutional neural network that combines multiple layers of convolution. Scaffolding helps identify the presence and location of cracks in concrete surface patterns.

This study suggests early detection of cracks in historic buildings. Citra captures images that do not show cracks and uses existing image processing algorithms to distinguish them from images that show cracks. The uniqueness of this study is that this study DCL-CNN model to help identify and classify crack types. Do these cracks appear in concrete or only on the surface of old buildings in Semarang city? Because of the insufficiency of observational methods for concrete structures, the study intends to make the following major contributions:

1) An effective and efficient classification framework with a combination of the number of convolution layers based on crack candidate areas is proposed to effectively categorize cracks and non-cracks. The more feature mapping that results from the number of convolution layers, the more detailed the system will be in detecting cracks in surface structures and deep fracture structures.

2) Comparison with other transfer learning methods and machine learning so that the CNN model produces the desired performance in classifying cracked and normal building locations

The image of cracks in ancient and historic buildings in Semarang, Indonesia, was used as data in this study. In this study, crack classification in old town buildings in Semarang requires several stages, as illustrated in Fig. 1. Images of old

building cracks are preprocessed. Preprocessing includes size adjustment, rotation, position translation, and flip processes.

The Citra image data set is split into a training data set and a validation (test) data set. Use preprocessed data to extract modal feature information from images using a pretrained model with transfer learning. This model is fed to a fully connected layer (FC) and trained after fusion. The top two layers of the FC layer contain 512 hidden units followed by the ReLU activation function. The final layer contains hidden units, followed by a sigmoidal activation function used for crack detection. Evaluate system performance using metrics such as Accuracy, Search Rate, Accuracy, F1 Score, MCC, and CK.



Fig. 1. Condition from the front side of the cracks of several ancient buildings in the old town of Semarang

### III. METHODOLOGY

The methodology for this paper shows in the Fig. 2. At the beginning, we need to process the crack image into through several levels of pre-processing, data augmentation, and training for each model. It is including modifications to the CNN model, retained model, and machine learning model, and all algorithms are tested using data testing

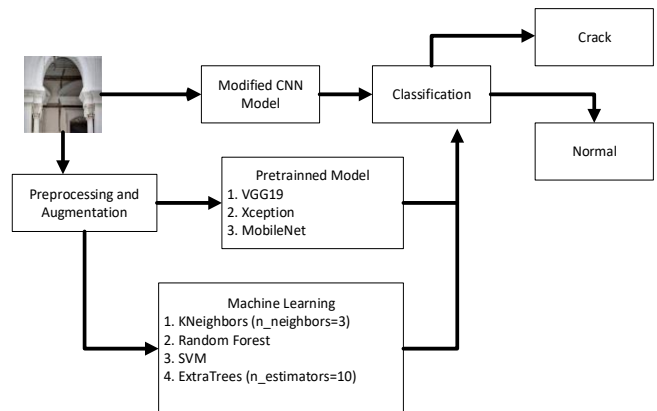


Fig. 2. Research methodology diagram



The experimental program is divided into four phases:

- creating a classified image dataset from primary data;
- developing the convolution layer from the CNN standard;
- comparing with machine learning and pre-trained models;
- running training experiments. The following sections provide information about each phase.

#### A. Image Data Set

The dataset employs for this study consists of 10,000 images with 512 x 512-pixel resolution. The image was taken from various concrete specimens after mechanical testing in the Semarang old town building. The main idea is to collect concrete service images from various surface views in order to diversify the data set and, as a result, the AI system that learns from this data set. The images were sliced into 224 x 224-pixel images to increase the data set without sacrificing resolution, resulting in a final data set of 10,000 samples, which were then manually classified into two categories: concrete surfaces with and without cracks. The dataset contains 5,000 images with cracks and 5,000 images without cracks. With a 70/30 split, the dataset is divided into training and validation datasets.

#### B. Convolution Layer from the CNN Standard

CNN1 is known as the base of the CNN method. For identified the crack building using Citra, this study used four variations of the CNN architecture. The usage of four variations of the CNN architecture used to find the best architecture for detecting the condition of cracks in ancient building structures in Semarang's old town. This study used GPU GTX 1650 RAM 2 x 8 GB 2400 MHz DDR4 for the computation.

Fig. 3 shows a diagram of the CNN1 architecture. This basic design can be extended to create CNN2 or CNN3 architectures. CNN3 is extensible to CNN4. CNN architectures are built to determine the impact of CONV layers and their activation functions on classification accuracy. Additionally, comparisons can be made between CNN1, CNN2, CNN3, and CNN4, or between design groups to determine if the best type of design is used for classifying old building crack images.

#### C. The Performance between Machine Learning and Pre-Trained Models

To evaluate classification performance, precision, recall, and accuracy matrices are used [16], [17]. To calculate the metric, add the sums of TP, FP, FN, and TN. True positive is represented by TP, false positive by FP, false negative by FN, and true negative by TN and properly measure ratios as positive detected elements, taking only positive predictions into account.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

The precision is stored in the denominator by FP; if it is high, the precision is low. However, the majority of the elements are predictably incorrect, and only a few are correct as positive, resulting in high precision values even if there are many FNs. As a result, a measure of the number of FN, namely recall, is required.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

and correctly measure ratios as positively detected elements, taking into account only elements with positive ground truth annotations. If the FN amount is large, the drawdown will be small and measures the ratio of correct predictions to all predictions.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

Accuracy can be used as a reliable summary metric for classification performance because the dataset is symmetric.

Table III displays the outcomes of all models in the primary dataset. To validate the model, we only use the test folder dataset. To compare performance, two statistical tests were performed: the Matthews correlation coefficient (MCC) and the Cohen's Kappa statistic [18-19]. In the case of unbalanced data sets, the Matthews correlation coefficient (MCC) is a popular performance metric. Despite the fact that the dataset used in this paper is balanced, it is defined by the mathematical equation number (4):

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}} \quad (4)$$

The MCC range is [1-1]. A MCC value closer to one is preferable. All of the models that were used performed admirably. The value is close to 1. In other words, the model correctly classifies the fracture image. Cohen's Kappa statistic is used to assess the degree of agreement between two raters who categorize objects into mutually exclusive groups, as shown mathematically in Equation (5).

$$CK = \frac{(p_o - p_e)}{(1 - p_e)} \quad (5)$$

In this case,  $p_o$  represents the rater's observations' relative agreement. The theoretical probability of random agreement is denoted by  $p_e$ . Using Equations (6)-(8), we can calculate  $p_o$  and  $p_e$  between raters (9).

$$p_o = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$P_e = \text{Probability of Positive} + \text{Propobability of Negative} \quad (7)$$

$$\text{Propability of Positive} = \frac{TP+FP}{TP+TN+FP+FN} \times \frac{TP+FN}{TP+TN+FP+FN} \quad (8)$$

$$\text{Propability of Negative} = \frac{FP+TP}{TP+TN+FP+FN} \times \frac{FN+TN}{TP+TN+FP+FN} \quad (9)$$

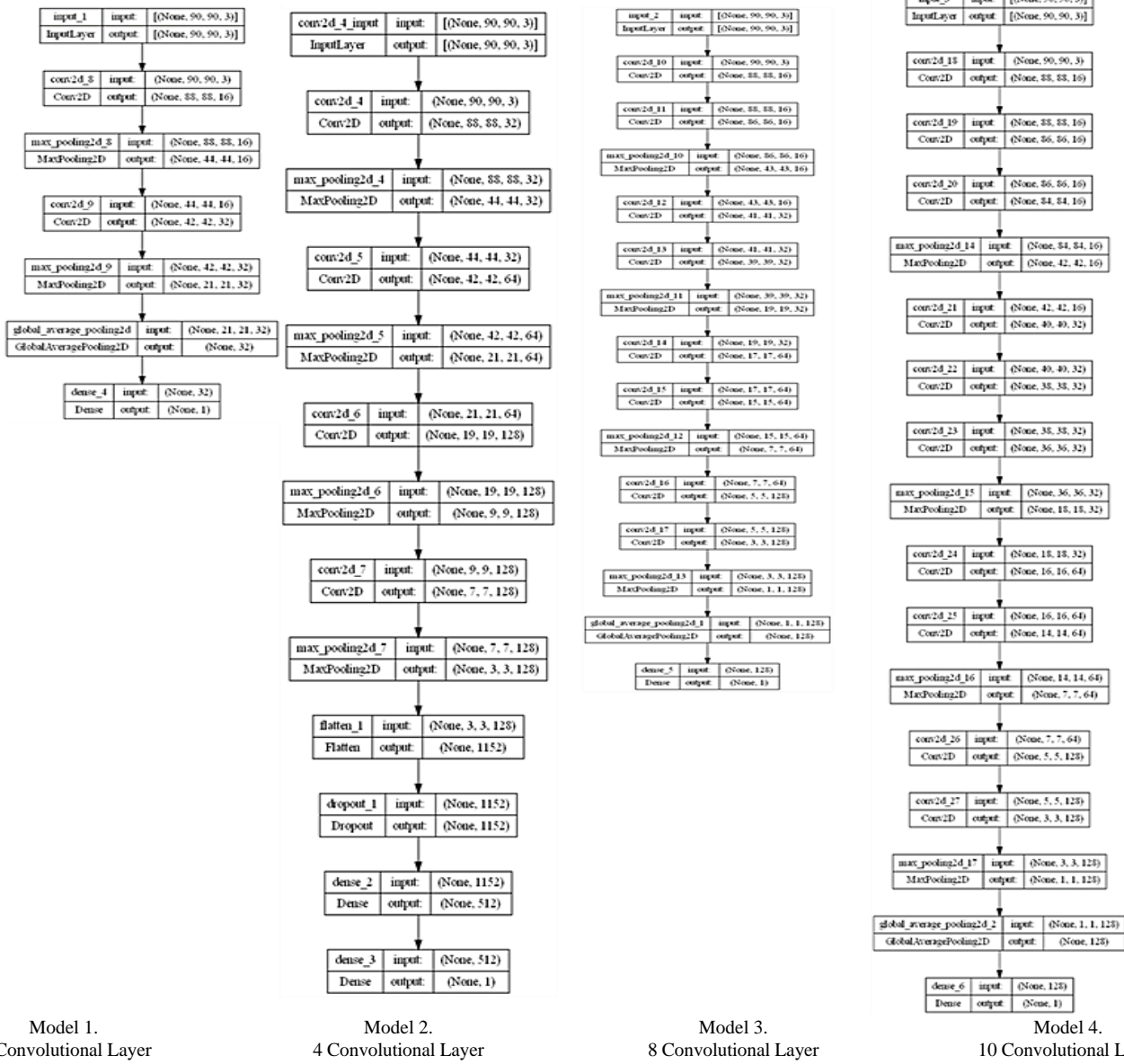


Fig. 3. Visualization for each CNN mode

Fig. 3 shows the architectural layout. CNN1 is the basis of the CNN method. This study employed four variations of the CNN architecture to determine the best architecture for detecting the presence or absence of cracks in ancient building structures in Semarang's old town. For the computation this study used GPU GTX 1650 RAM 2 x 8 GB 2400 MHz DDR4.

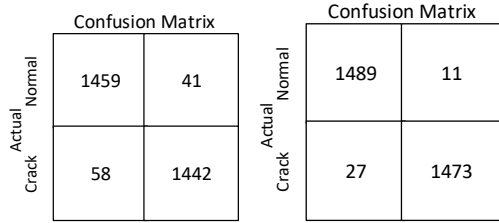
This basic design can be extended to create CNN2 or CNN3 architectures, as shown in the CNN1 architecture diagram in Fig. 3. CNN3 is extensible to CNN4. CNN architectures are built to determine the impact of CONV layers and their activation functions on classification accuracy. In addition, comparisons can be made between CNN1, CNN2, CNN3, and CNN4 or between design groups to determine if the most appropriate type of design is being used for classifying old building crack images.

IV. EXPERIMENT AND DISCUSSION

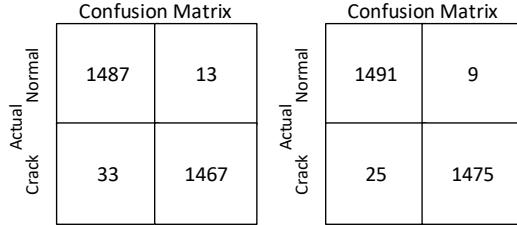
Fig. 4(a) shows Model 1 with two convolution layers correctly predicts 1459 (TP) normal and 1442 (TN) cracked images, while 41 (FN) normal images are predicted to be cracked and 58 (FP) cracked images are predicted to be normal

Fig. 4(b) shows Model 2 with four convolution layers correctly predicts 1489 (TP) normal and 1473 (TN) cracked images, while 11 (FN) normal images are predicted to be cracked and 27 (FP) cracked images are predicted to be normal. Fig. 4(c) shows Model 3 with 8 convolution layers correctly predicts 1487 (TP) normal and 1467 (TN) cracked images, as well as 13 (FN) normal images that are cracked and 33 (FP) cracked images that are normal. Fig. 4(d) shows Model 4 with 10 convolution layers correctly predicts 1491 (TP) normal and 1475 (TN) crack images, as well as 9 (FN) normal images that are cracked and 25 (FP) cracked images that are

normal. The FN on Model 1.d is smaller than on the other three models.



(a) Model 1.2 Convolutional Layer, (b). Model 2.4 Convolutional Layer



(c) Model 3.8 Convolutional Layer, (d). Model 4.10 Convolutional Layer

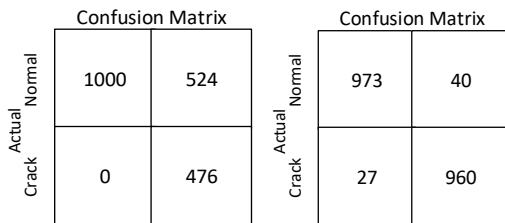
Fig. 4. Confusion matrix results for each CNN modification model

By identifying knowing TP, TN, FN, FP, performance values it generates such as accuracy, precision, recall, F1, MCC and CK values as shown in Table I, Table II and Table III. Table I shows the performance of Fig. 4 Confusion Matrix Results for each CNN modification model

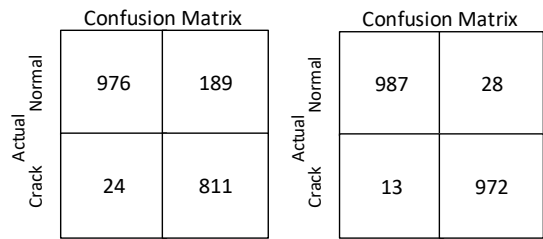
TABLE I. PERFORMANCE COMPARISON OF DIFFERENT CNN MODIFICATION METHODS

N_LAYER CONVOLUTIONAL	ACCURACY SCORE	RECALL SCORE	PRECISION SCORE	F1 SCORE	MCC (%)	CK (%)
2 LAYER	96.70	97.23	96.13	96.68	93.41	96.69
4 LAYER	98.73	99.26	98.20	98.73	97.47	98.73
8 LAYER	98.43	98.34	98.53	98.43	96.94	98.46
10 LAYER	98.87	99.40	98.33	98.86	97.74	98.86

Table I displays that the convolution layer design with 10 convolution layers has the highest F1 score of 98.86% and 98.87% accuracy among other models, as well as 98.33% precision and 99.40% recall. The Deca Convolutiol Layer Neural Network (DCL-NN) outperforms the other models in the popular statistical tests MCC and CK (Cohen's Kappa), with values of 97.74% and 98.86%.



(a). KNeighbors (n\_neighbors=3) (b). Random Forest



(c) SVM (d) ExtraTrees (n\_estimators=10)

Fig. 5. Confusion matrix results for each machine learning model

In Fig. 5(a), the KNN model correctly predicts 1000 (TP) normal and 476 (TN) crack images, while 524 (FN) normal images are predicted as cracked and 0 (FP) cracked images are predicted as normal. In Fig. 5(b), the Random Forest model correctly predicts 973 (TP) normal and 960 (TN) cracked images, with 40 (FN) normal images predicted as cracked and 27 (FP) cracked images predicted as normal. In Fig. 5(c), the SVM model correctly predicts 976 (TP) normal and 811 (TN) cracked images, with 189 (FN) normal images predicted to be cracked and 13 (FP) cracked images predicted to be normal. In Fig. 5(d), the Extra Tree model correctly predicts 987 (TP) normal and 972 (TN) cracked images, with 28 (FN) normal images predicted as cracked and 13 (FP) cracked images predicted as normal. When compared to the other three machine learning models, the Extra Trees model has a lower FN. Table II shows the performance of Fig. 5. Confusion Matrix Results for each machine learning model

TABLE II. PERFORMANCE COMPARISON OF DIFFERENT MACHINE LEARNING METHODS

Machine Learning	Accuracy Score	Recall Score	Precision Score	F1 Score	MCC (%)	CK (%)
K Neighbors (n_neighbors=3)	74	83	74	72	55.89	69.85
Random Forest	97	97	97	97	93.31	96.64
SVM	89	90	89	90	79.79	88.89
ExtraTrees (n_estimators=10)	98	98	98	98	95.91	97.94

According to Table II, the ExtraTrees machine learning model has the highest F1 score of 98% and 98% accuracy among other models, as well as 98% precision and 98% recall. In the case of popular statistical tests such as MCC and CK (Cohen's Kappa), the ExtraTrees machine learning model outperforms the other models with values of 95.91% and 97.94%, respectively.

Fig. 6(a) illustrates the transfer learning model VGG19 correctly predicts 1487 (TP) normal and 1470 (TN) cracked images, with 13 (FN) normal images predicted to be cracked and 30 (FP) cracked images predicted to be normal. As for the Fig. 6(b) shows the Xception model correctly predicts 1444 (TP) normal and 1497 (TN) cracked images, with 56 (FN) normal images predicted as cracked and 3 (FP) cracked images predicted as normal. Fig. 6(c) illustrates the MobileNet model correctly predicts 1472 (TP) normal and 1434 (TN) cracked images, with 28 (FN) normal images predicted as cracked and

66 (FP) cracked images predicted as normal. The VGG19 model has a lower FN than the other three transfer learning models. The performance of Fig. 6, Matric Confusion Results for each model of Pre-Trained CNN is shown in Table III.

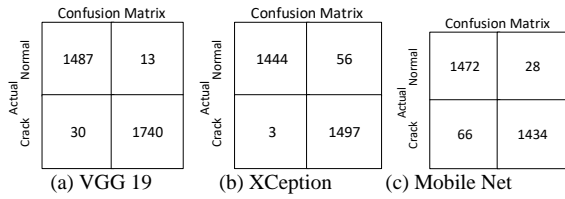


Fig. 6. Matric confusion results for each model of pre-trained CNN

TABLE III. PERFORMANCE COMPARISON OF DIFFERENT TRANSFER LEARNING METHODS

Transfer Learning	Accuracy Score	Recall Score	Precision Score	F1_Score	MCC (%)	CK (%)
VGG19	98.57	99.12	98.00	98.56	97.14	98.56
Xception	98.03	96.39	99.80	98.07	96.13	98.05
MobileNet	96.87	98.08	95.60	96.83	93.76	96.85

Table III shows that the VGG19 model transfer learning design had the highest F1 score of 98.56% and 98.57% accuracy, as well as 98.00% precision and 99.12% recall among other models. The transfer learning performance of the VGG19 model outperforms the other models in popular statistical tests such as MCC and CK (Cohen's Kappa), with values of 97.14% and 98.56%, respectively.

In Model 1, after 10 iterations, the program automatically stops training, and the correct model rate is approximately 97.67%. (Loss 0.1141). As illustrated in Fig. 9, the maximum true rate is reached after the tenth iteration. It takes about 13 seconds. As illustrated in Fig. 7, the correct rate gradually stabilizes in later training stages due to the continuous reduction of the learning rate. In Model 2, after 10 iterations, the program automatically stops training, and the correct model rate is approximately 98.57% (0.0646).

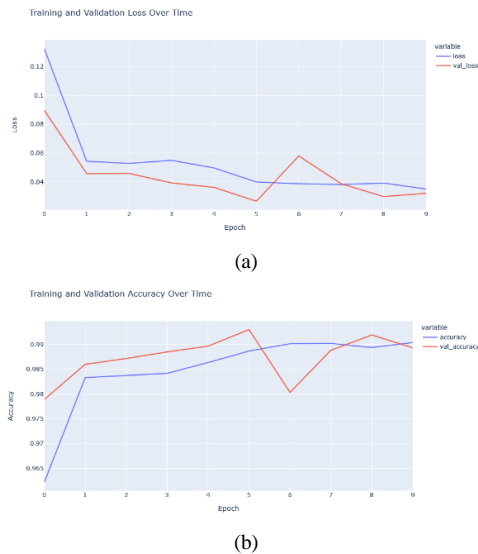


Fig. 7. Graph of accuracy and loss in model 1 with 2 convolutional layer

Fig. 8 shows that the maximum true rate is reached after the tenth iteration. It takes about 28 seconds. As shown in Fig. 8, the correct rate gradually stabilizes in later training stages due to the continuous reduction of the learning rate. In Model 2, after 10 iterations, the program automatically stops training, and the correct model rate is approximately 98.90% (Loss 0.0561).

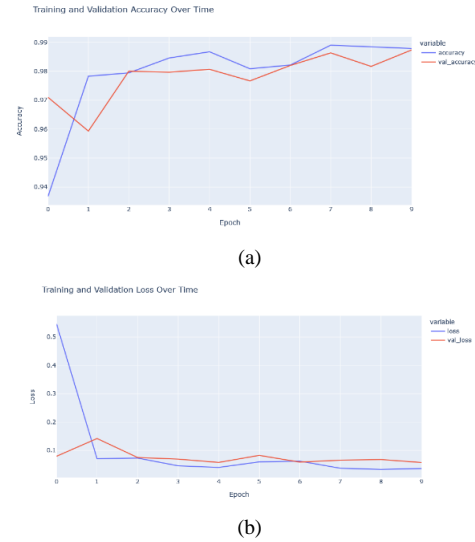


Fig. 8. Graph of accuracy and loss in model 2 with 4 convolutional layer

Figure 9 shows that the maximum true rate is reached after the tenth iteration. It takes about 31 seconds. As illustrated in Figure 9, the correct rate gradually stabilizes in later training stages due to the continuous reduction of the learning rate. In Model 4, after 10 iterations, the program automatically stops training, and the correct model rate is approximately 99.17%. (Loss 0.0284)

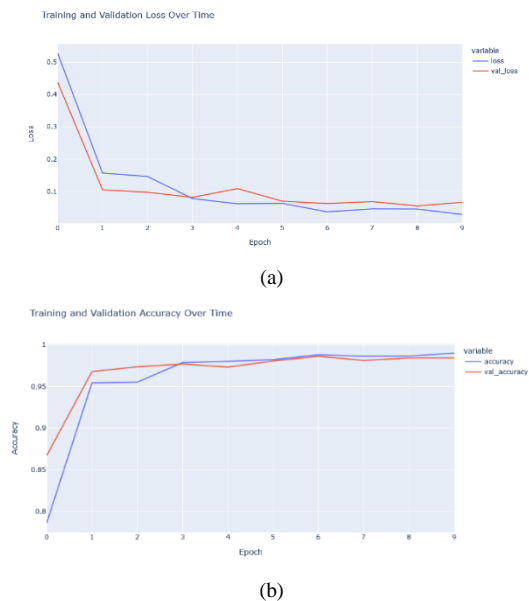


Fig. 9. Graph of accuracy and loss in model 2 with 8 convolutional layer

Fig. 12 shows that the maximum true rate is reached after the tenth iteration. It takes about 32 seconds. As illustrated in Fig. 12, the correct rate gradually stabilizes in later training stages due to the continuous reduction of the learning rate.

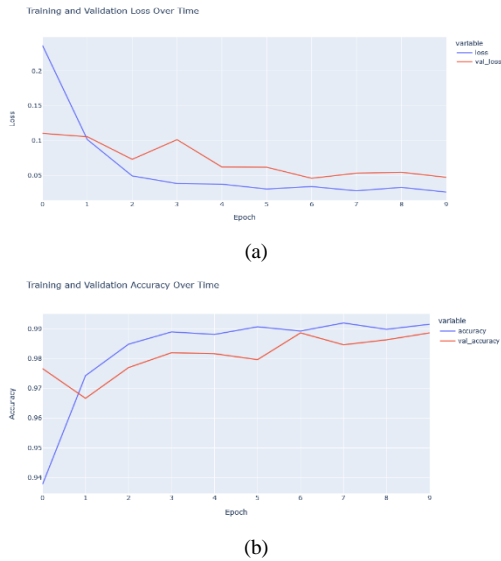


Fig. 10. Graph of accuracy and loss in model 2 with 10 convolutional layer

Fig. 10 shows that the maximum true rate is reached after the tenth iteration. It takes about 32 seconds. As illustrated in Fig. 10, the correct rate gradually stabilizes in later training stages due to the continuous reduction of the learning rate.

The CNN model's performance is also compared to that of other transfer learning models, such as VGG19, Exception, and MobilNet. The following performance graphs are generated for each transfer learning model:

The VGG model automatically stops training after 10 iterations, and the correct model rate is approximately 98.93%. (Loss 0.0384). As shown in Fig. 13, the maximum true rate is reached after the tenth iteration. It takes approximately 61 seconds. As shown in Fig. 13, the correct rate gradually stabilizes in later training stages due to the continuous reduction of the learning rate. In Model Exception, the program automatically stops training after 10 iterations, and the correct model rate is approximately 98.67% (Loss 0.0526).

As shown in Fig. 11, the maximum true rate is reached after the tenth iteration. It takes approximately 61 seconds. As shown in Fig. 11, the correct rate gradually stabilizes in later training stages due to the continuous reduction of the learning rate. In Model Exception, the program automatically stops training after 10 iterations, and the correct model rate is approximately 98.67% (Loss 0.0526).

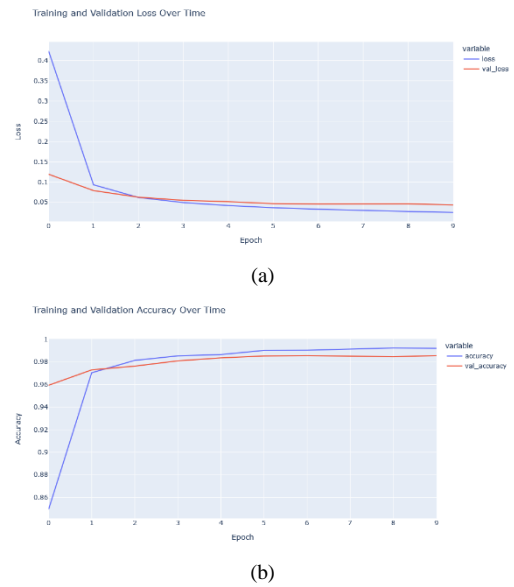


Fig. 11. VGG19 model accuracy and loss graphs

As illustrated in Fig. 12, the maximum true rate is reached after the tenth iteration. It takes about 63 seconds. As illustrated in Fig. 12, the correct rate gradually stabilizes in later training stages due to the continuous reduction of the learning rate. The MobileNet model, after 10 iterations, the program automatically stops training, and the correct model rate is approximately 96.67% (Loss 0.0993).

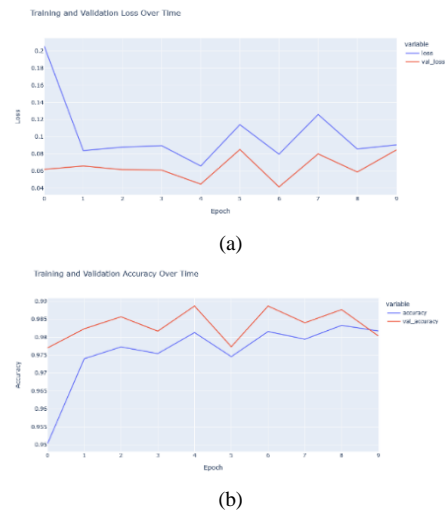


Fig. 12. Graph of accuracy and loss in exception model

As shown in Fig. 13, the maximum true rate is reached after the tenth iteration. It takes approximately 30 seconds. As shown in Fig. 13, the correct rate gradually stabilizes in later training stages due to the continuous reduction of the learning rate.

## V. CONCLUSION

This research contributes to crack classification at historical building by using four architectural design variants of convolutional neural networks (CNNs), namely, CNN1-CNN4. The research methodology used in this study is an experimental cracking classification using four variations of convolutional neural network (CNN) architecture design (CNN1 - CNN4).

Experimental results show that CNN4 (Deca Convolutional Layer Neural Network/DCL-NN) provides the best classification results for concrete cracks in old buildings compared to other architectural designs tested. The DCL-NN architecture has an accuracy of 98.87%, precision of 99.40%, recall of 98.33%, F1 score of 98.86%, MCC of 97.74%, and CK of 98.86%.

In addition, results from comparisons with pre-trained CNN algorithm methods such as VGG19, Xception, MobileNet, and machine algorithms such as KNeighbors (n neighbors=3), Random Forests, Support Vector Machines (SVM), ExtraTrees (n estimators) Classification using learning. = 10), he shows the superiority of DCL-NN in classifying concrete cracks in old buildings in Semarang city.

## ACKNOWLEDGMENT

This research supported by Research grant 2022 (Research Publication International /RPI) from Universitas Diponegoro SK. Rector Universitas Diponegoro No: 215 /UN7.A/HK/VII/2022; No.SPK:569 108/UN7.D2/PP/VII/2022.

## REFERENCES

- h. werdiningsih s. r. sari, a. r. harani, "Pelestarian dan Pengembangan Kawasan Old Town sebagai landasan budaya kota Semarang." Modul, vol. 17, no. 1, pp. 49–55, 2017.
- H. N. Agastya Grahadwiswara, Zaenal Hidayat, "Pengelolaan Kawasan Old town Semarang Sebagai Salah Satu Kawasan Pariwisata Di Kota Semarang," J. Public Policy Manag. Rev., vol. 3, no. 4, 2014.
- D. G. Aggelis, N. Alver, and H. K. Chai, "Health monitoring of civil infrastructure and materials," Sci. World J., vol. 2014, 2014, doi: 10.1155/2014/435238.
- M. Gavilan et al., "Adaptive road crack detection system by pavement classification," Sensors, vol. 11, no. 10, pp. 9628–9657, 2011, doi: 10.3390/s111009628.
- Y. C. Tsai, V. Kaul, and R. M. Mersereau, "Critical assessment of pavement distress segmentation methods," J. Transp. Eng., vol. 136, no. 1, pp. 11–19, 2010, doi: 10.1061/(ASCE)TE.1943-5436.0000051.
- A. Hafid and D. S. Rahayu, "Pengukuran Kedalaman Retak pada Beton Menggunakan UPVT," Sigma Epsilon, vol. 24, no. 1, pp. 1–8, 2020.
- P. Wang, Y. Hu, Y. Dai, and M. Tian, "Asphalt Pavement Pothole Detection and Segmentation Based on Wavelet Energy Field," Math. Probl. Eng., vol. 2017, 2017, doi: 10.1155/2017/1604130.
- T. Yamaguchi, S. Nakamura, R. Saegusa, and S. Hashimoto, "Image-based crack detection for real concrete surfaces," IEEE Trans. Electr. Electron. Eng., vol. 3, no. 1, pp. 128–135, 2008, doi: 10.1002/tee.20244.
- Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
- J. Zhu and J. Song, "An intelligent classification model for surface defects on cement concrete bridges," Appl. Sci., vol. 10, no. 3, 2020, doi: 10.3390/app10030972.
- M. A. Moni and F. Hasan, "CNN Based on Transfer Learning Models Using Data Augmentation and CNN Based on Transfer Learning Models Using Data Augmentation and Transformation for Detection of Concrete Crack," Algorithms, no. August, pp. 2–17, 2022, doi: 10.3390/a15080287.

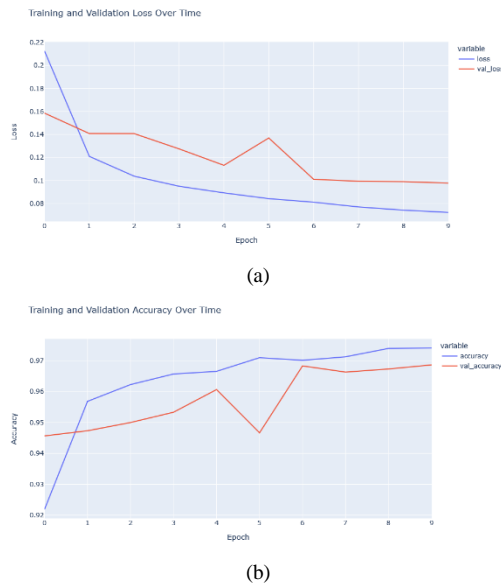


Fig. 13. Accuracy and loss graphs on the MobileNet model

Researchers in this study conducted research that was also compared to other researchers using the same dataset. Zhang et al. [20] classified cracks using CNN and four convolution layers, two of which were fully connected. ConvNets is the name of the method used. Precision and recall rates are 86.96% and 92.51%, respectively. In classifying aligned images, Fang et al. [21] employ three convolution layers and three Fully Connected Layers. Precision and recall rates for experiments using the same data as the researchers were 18.4% and 94.3%, respectively. While the researchers used 10 convolution layers and one Fully Connected layer to achieve a recall and precision of 99.40% and 98.33%, respectively. Our model outperforms both works [1] and [2], as shown in the Table IV below.

TABLE IV. COMPARISON OF PERFORMA WITH OTHER RESEARCHERS

Other Researchers	Number of Convolution Layers	Number of Fully Connected Layers	Precision	Recall
Zhang et al. [20]	4	2	86.96%	92.51%
Fang et al. [21]	3	2	18.40%	94.30%
Proposed Model	10	1	98.33%	99.40%

Each neuron in the convolution layer needs to be transformed into one-dimensional data. First before it can be included in a fully-connected layer. Also, because it causes the data to lose its spatial information and is not reversible, while the fully connected layer can only be implemented at the end of the network CNN [22].

Prior researchers used 2 Fully Connected Layers (FCL) to prevent loss of image spatial information and a long duration so that classification performance can be maintained, but with this study, we used a large number of convolution layers and a low number of FCL, to produce high performance.

- [12] Z. Y. J. Zhang L., Yang F., Zhang Y.D., "Performance Evaluation of Different Algorithms for Crack Detection in Concrete Structures," in Proceedings of the International Conference on Image Processing, ICIP, 2016, pp. 6–11.
- [13] W. Wang and C. Su, "Automation in Construction Semi-supervised semantic segmentation network for surface crack detection," *Autom. Constr.*, vol. 128, no. April, p. 103786, 2021, doi: 10.1016/j.autcon.2021.103786.
- [14] Z. Fan, C. Li, Y. Chen, P. Di Mascio, X. Chen, and G. Zhu, "Ensemble of Deep Convolutional Neural Networks for Automatic Pavement Crack Detection and Measurement," *Coatings*, pp. 1–15, 2020, doi: 10.3390/coatings10020152.
- [15] H. Xu, X. Su, Y. Wang, H. Cai, K. Cui, and X. Chen, "applied sciences Automatic Bridge Crack Detection Using a Convolutional Neural Network," *Appl. Sci.*, pp. 1–14, 2019.
- [16] A. K. Nugroho and M. H. Purnomo, "Quad Convolutional Layers ( QCL ) CNN Approach for Classification of Brain Stroke in Diffusion Weighted ( DW ) - Magnetic Resonance Images ( MRI )," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 1, pp. 414–427, 2022, doi: 10.22266/ijies2022.0228.38.
- [17] A. K. Nugroho, "Utilizing the Hepta Convolutional Layer Neural Network ( HCL-NN ) Based on a Multi Optimizer for the Classification of Brain Stroke MRI," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 3, pp. 304–318, 2022, doi: 10.22266/ijies2022.0630.26.
- [18] D. Chicco, M. J. Warrens, and G. Jurman, "The Matthews correlation coefficient ( MCC ) is more informative than Cohen ' s Kappa and Brier score in binary classification assessment," *IEEE Access*, vol. X, pp. 1–15, 2021, doi: 10.1109/ACCESS.2021.3084050.
- [19] N. Hasdyna, "Analisis Matthew Correlation Coefficient pada K-Nearest Neighbor dalam Klasifikasi Ikan Hias," *Inform. J.*, vol. 5, no. 2, pp. 57–64, 2020.
- [20] Y. J. Z. Lei Zhang , Fan Yang , Yimin Daniel Zhang, "ROAD CRACK DETECTION USING DEEP CONVOLUTIONAL NEURAL NETWORK," in 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 3708–3712, doi: 10.1109/ICIP.2016.7533052.
- [21] F. Fang, L. Li, Y. Gu, H. Zhu, and J. Lim, "A novel hybrid approach for crack detection," *Pattern Recognit.*, vol. 107, 2020, doi: 10.1016/j.patcog.2020.107474.
- [22] R. O.aditya yanuar (2018) Fully-connected layer CNN Dan Implementasinya, Universitas Gadjah Mada. Available at: <https://machinelearning.mipa.ugm.ac.id/2018/06/25/fully-connected-layer-cnn-dan-implementasinya/> (Accessed: January 21, 2023).

# Convolutional Transformer based Local and Global Feature Learning for Speech Enhancement

Chaitanya Jannu<sup>1</sup>, Sunny Dayal Vanambathina<sup>2</sup>

School of Electronics Engineering, VIT-AP University, Amaravati, India<sup>1,2</sup>

**Abstract**—Speech enhancement (SE) is an important method for improving speech quality and intelligibility in noisy environments where received speech is severely distorted by noise. An efficient speech enhancement system relies on accurately modelling the long-term dependencies of noisy speech. Deep learning has greatly benefited by the use of transformers where long-term dependencies can be modelled more efficiently with multi-head attention (MHA) by using sequence similarity. Transformers frequently outperform recurrent neural network (RNN) and convolutional neural network (CNN) models in many tasks while utilizing parallel processing. In this paper we proposed a two-stage convolutional transformer for speech enhancement in time domain. The transformer considers global information as well as parallel computing, resulting in a reduction of long-term noise. In the proposed work unlike two-stage transformer neural network (TSTNN) different transformer structures for intra and inter transformers are used for extracting the local as well as global features of noisy speech. Moreover, a CNN module is added to the transformer so that short-term noise can be reduced more effectively, based on the ability of CNN to extract local information. The experimental findings demonstrate that the proposed model outperformed the other existing models in terms of STOI (short-time objective intelligibility), and PESQ (perceptual evaluation of the speech quality).

**Keywords**—Convolutional neural network; recurrent neural network; speech enhancement; multi-head attention; two-stage convolutional transformer; feed-forward network

## I. INTRODUCTION

In the area of speech processing, speech enhancement is crucial. The main objective is to enhance speech that has been impaired by background noise in terms of clarity and quality. Numerous applications including powerful speech recognition, teleconferencing, and hearing aids, employ it as a pre-processor. Recent advances in deep learning have made it possible to develop several data driven methods to solve traditional estimation problems without having to rely on supervision. The majority of existing deep learning models for speech enhancement, such as convolutional neural network (CNN) and recurrent neural network (RNN), are implemented in the time-frequency (T-F) domain. Those methods use short-time Fourier transforms (STFT) to train on spectral magnitude. To reconstruct the time-domain signal using the inverse short-time Fourier transform (iSTFT), it is necessary to take into account the phase of noisy speech along with the improved speech magnitude. Despite some impressive results [1-4], T-F domain methods still have two key drawbacks. In the first place, Fourier transforms add an additional overhead to fast speech denoising. Second, during the denoising

process, the noisy phase is generally ignored. Nevertheless, phase information has been shown to be significant for improving speech quality [5]. To achieve better enhancement results, there are some studies that look at magnitude as well as phase simultaneously during training [6]. Several recent works have directly estimated the clean speech in time domain from noisy raw data [7-11].

## II. RELATED WORK

A fully convolutional network (FCN) has been proposed by Fu et al. [12] for raw waveform-based speech enhancement which improves the quality compared to masking-based methods [13-14]. In distinction to both masking and mapping techniques that use noisy phase to reconstruct enhanced speech, FCN provides speech enhancement by simply mapping from a noisy speech to the matching clean speech. Many studies have looked into CNN or RNN-based encoder-decoder frameworks. CNN requires more convolutional layers to enlarge the receptive field when modelling long-range sequences such as speech. The Convolutional recurrent networks (CRNs) [15] are used to extract long-context information using CNN's which are familiar for feature extraction and RNNs' which are familiar for temporal modelling. For processing long-term temporal sequences, a dilated convolutional neural network has been proposed [16]. K. Tan et al. [17] proposed GRN with Dilated convolutions for supervised SE. The convolutions with dilation enlarge the receptive field without sacrificing resolution compared to regular convolutions by maintaining the same kernel size and network depth. At various SNR levels the GRN exhibits higher generalization ability to untrained speakers when compared to the LSTM model in [18]. A CNN model in time domain is presented by considering frequency domain loss to enhance the quality of corrupted speech [19]. Although the work in [19] can achieve cutting-edge performance, the issue of real-time enhancement is not addressed. Later the same authors proposed densely connected network [20] for real time SE and achieved better performance than [19].

The RNNs such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) are usually used to model long-term sequences with order information. Although introducing temporal convolutional network (TCN) blocks [10] or LSTM layers between encoder and decoder can improve denoising performance by extracting higher-level features and expanding receptive fields [11], the contextual information of speech is generally ignored, limiting denoising performance. Recently Li, Qinglong, et al. [21] proposed a DCTCRN model for speech enhancement where discrete cosine transform is used as input to the model for processing noisy speech and uses



LSTM as bottleneck. Even though the performance of the model is better, the LSTMs are easily prone to the problem of overfitting and it also requires large time to train and are also sensitive to random weight initializations. Kai Zhen et al. proposed a DCCRN [22] network for SE in time domain where densenets and GRUs are used for context aggregation. This model yields better intelligibility scores compared to convolutional base line models but quality is less compared to Wave-U-Net model [7]. Recently Lin et al. proposed a SASE model [23] with self-attention for speech enhancement which extracts speaker information using LSTM layers and also CRN model is used for speech enhancement based on embedding of speaker information which performs better than CRNs and LSTM models but the quality is not satisfactory. A unique ARN for time-domain speech improvement was recently proposed by Pandey A., et al. [24] in order to promote cross-corpus generalization. Self-attention and feedforward blocks are added to RNN to create ARN. It has been shown that transformer neural networks are capable of resolving the long-dependency problem effectively and are capable of operating in parallel, so they are good at tackling a wide range of natural language processing tasks [25]. As the transformer only uses the attention mechanism, the vanishing gradient and exploding gradient problem of RNNs are also solved. A transformer model for speech enhancement was proposed in [26], which has a comparatively large size of model. Transformer enhances the training speed and prediction accuracy compared to RNNs [25]. By combining the advantages of LSTM and multi-head attention mechanism, Yu et al. [27] proposed SETransformer. When compared with a standard transformer and an LSTM model, SETransformer showed better denoising performance. Recently Wang et al. [28] introduced an end-to-end speech enhancement two-stage transformer neural network (TSTNN).

The limitations of existing frameworks are the CNN models [12,19] requires more convolutional layers to enlarge the receptive field when modelling long-range sequences such as speech. The disadvantage of RNN-based models [18] is that they cannot perform parallel processing, resulting in high computation complexity. The limitation of TCN [10] is that the contextual information of speech is generally ignored, limiting denoising performance. The CRN model [21] suffers with computational complexity. The transformer models present in [26], [27] performs better than convolutional and RNN baselines but they are implemented in the time-frequency (T-F) domain. To achieve better enhancement results, there are some studies that look at magnitude as well as phase simultaneously during training. Recently Wang et al. [28] introduced an end-to-end speech enhancement two-stage transformer neural network (TSTNN) which performs better than [26] and [27]. But in TSTNN model same structure of FFN is used for extracting local and global features. And also, they have not concentrated on the local (short-term) noises present at the output of encoder.

Motivated by recent success of transformer neural networks in natural language processing tasks [25] and SE [26-28] we propose a novel two-stage convolutional transformer neural network that enhances monaural speech in a time domain. The transformers are capable of resolving the

long-range dependency problem effectively and are capable of operating in parallel. As the transformer only uses the attention mechanism, the vanishing gradient and exploding gradient problem of RNNs are also solved. In this work, we propose a two-stage convolutional transformer neural network that enhances monaural speech in a time domain from end to end which differs from the existing transformer models such as T-gsa [26], SETransformer [27] and two-stage transformer neural network (TSTNN) [28]. Based on the transformer's ability to model sequences and the dual-path network's ability to extract contextual information [29], we propose a two-stage convolutional transformer neural network.

In this paper the proposed transformer considers global information as well as parallel computing, resulting in a reduction of long-term noise. Unlike the transformer proposed in [28], we proposed a novel transformer structure for intra and inter transformers to extract the local as well as global features of noisy speech using CNN and RNN layers [30]. A 1D-Conv layer and temporal convolution module (TCM) are used in intra transformer where local features are extracted and a 1D-Conv layer and Bi-directional long short-term memory (Bi-LSTM) are used in inter transformer where global features are extracted.

Moreover, in the proposed model, a CNN module is also added to intra and inter transformers so that short-term noise can be reduced more effectively, based on the ability of CNN to extract local information. The proposed model enhances the local information modelling capability of the traditional transformer model by adding a convolution layer. The convolutional module consists of Depth and Point wise convolutions instead of normal convolution to increase the speed of operation of model and both of them have less parameters compared to normal convolution. The details of convolutional module and modified layers of feed forward network are explained in the following sections.

Our contributions:

- In contrast to RNNs and CNNs, the proposed transformer model uses parallel processing and the long-term dependencies can be modelled more efficiently with multi-head attention (MHA) by using sequence similarity.
- The novelty of the proposed work is different transformer structures for intra and inter transformers are used for extracting the local as well as global features of noisy speech using CNN and RNN layers. In the proposed model a CNN module is also added to intra and inter transformers so that short-term noise can be reduced more effectively, based on the ability of CNN to extract local information.
- The proposed model enhances the local information modelling capability of the traditional transformer model by adding a CNN module.

The remainder of this work is structured as follows. The presents the related works. Section II presents related works and we explained the details of proposed two-stage convolutional transformer in Section III. Section IV presents

the experimental findings. Section V presents discussion and the paper is concluded in Section VI.

### III. ARCHITECTURE OF PROPOSED TWO-STAGE CONVOLUTIONAL TRANSFORMER

The overall architecture of proposed model is shown in Fig. 1. The model consists of four modules: Encoder, Transformer module, Masking module and Decoder. There are two convolution layers in the encoder among them the first one is to increase the channels to 64 and the second one is

used to halve the size of the frame. The transformer module internally consists of four transformer blocks responsible for feature extraction. The detailed transformer module is explained in following Section 3.C. The masking module is used to obtain the mask using 2-way convolution and nonlinear activations. In the decoding phase the reconstruction of features is done by using dilated dense block and sub-pixel convolution. Finally, we will get the enhanced speech after normalization and PReLU.

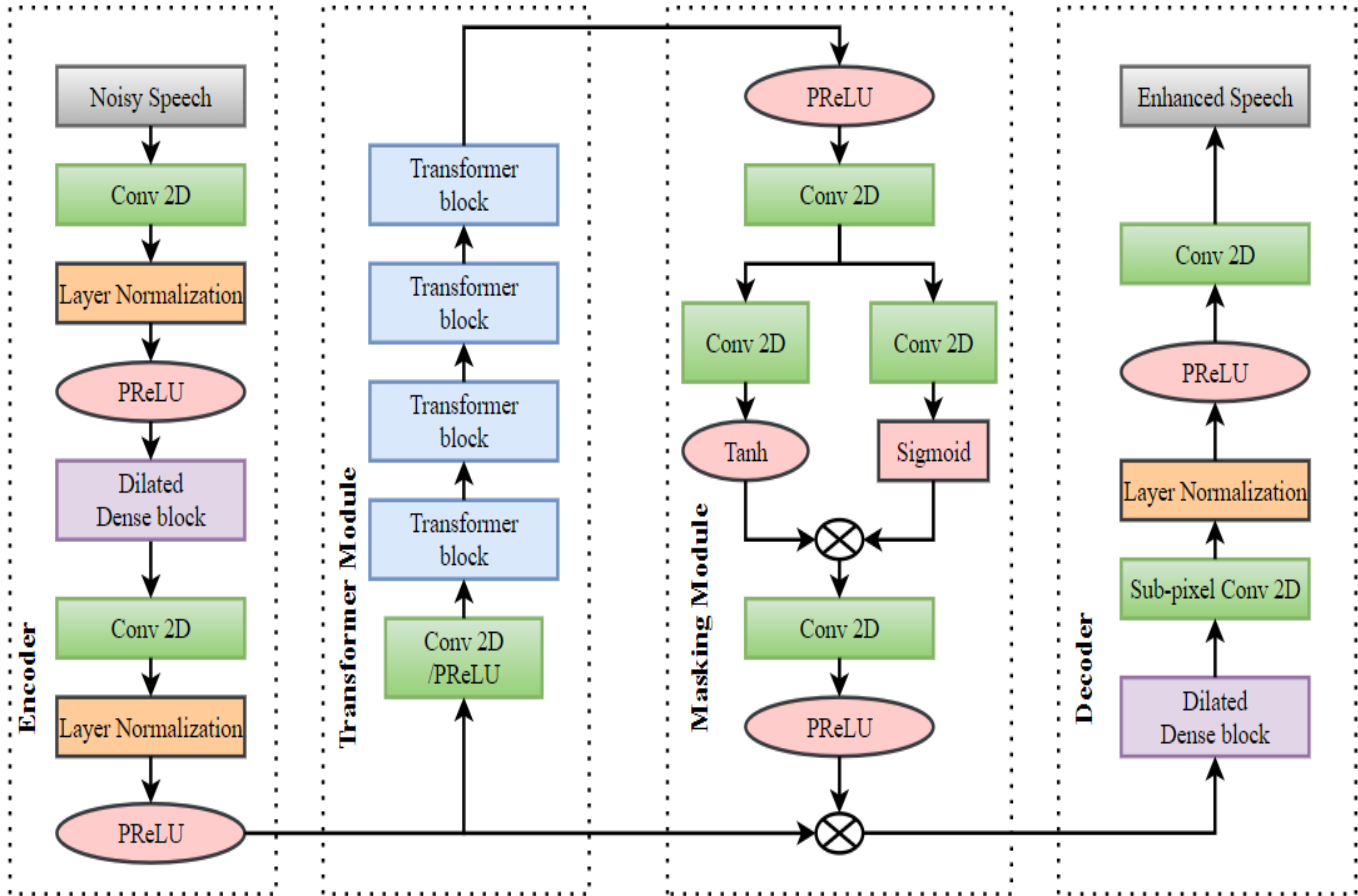


Fig. 1. Architecture of proposed convolutional transformer

#### A. Segmentation and Overlap-Add

In this stage the input noisy speech  $Y \in R^{1 \times L}$  is splitted into frames with length of each frame as  $F$  and hop size as  $H$ . The 3D tensor  $I \in R^{1 \times N \times F}$  is then created by stacking all frames as input to the encoder. The length of input noisy speech is denoted by  $L$  and total frame count is denoted by  $N$ .

$$N = [(L - F)/(F - H) + 1] \quad (1)$$

To recover the enhanced speech waveform the overlap-add operation is used at decoder.

#### B. Encoder

There are two convolution layers in the encoder among them the first one is to increase the channels to 64 with filter

size of (1,1) and the second one is used to halve the frame size with filter and stride of (1,3) and (1,2). And a 4-layer dilated dense block [20] is inserted in between them. The Layer normalization and PReLU nonlinearity [30-31] are applied to all convolutional layers.

#### C. Transformer Module

The transformer module internally consists of four transformer blocks stacked together where feature extraction is performed to extract local as well as global features. Before giving the output of encoder to input of transformer, we use convolution with a kernel of size (1, 1) followed by PReLU activation to halve the channel dimension to reduce the computational complexity of the transformer network. The transformer block is shown in below Fig. 2.

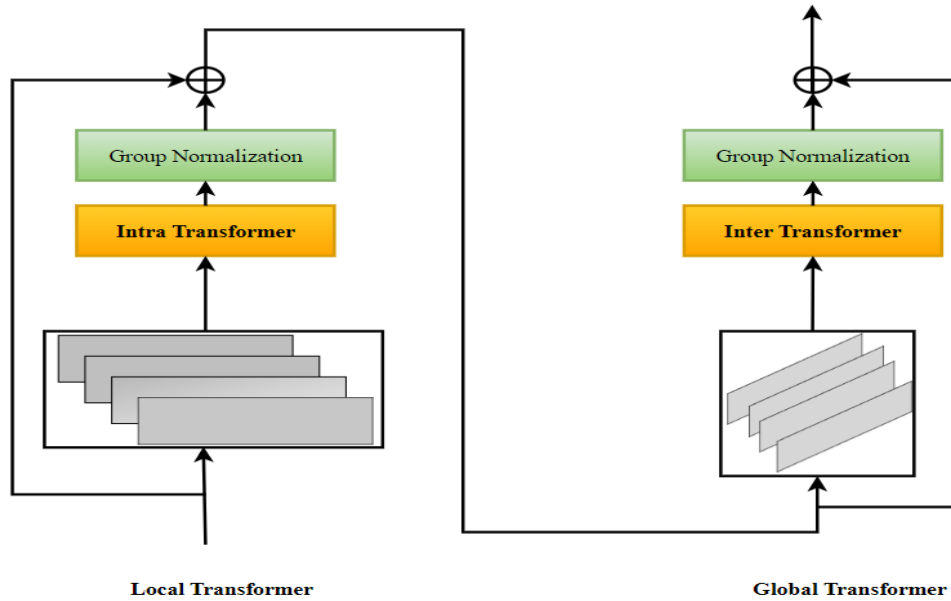


Fig. 2. Transformer block

1) *Improved transformer*: The transformer block consists of intra and inter transformers as shown in above Fig. 2. To independently model the local chunks intra transformer is used and inter transformer is used to extract global dependencies by summarizing information of all blocks.

Three important modules are included in the original transformer encoder [25]: positional encoding, multi-head

attention, and position-wise feed-forward. Due to the fact that the positional encoding is not appropriate for acoustic sequences the authors of baseline model [28] designed GRUs to learn positional information by replacing the first fully connected layer of feed-forward networks with RNNs to track order information [32-33]. The structure of feed forward network used in baseline model [28] is shown in Fig. 3.

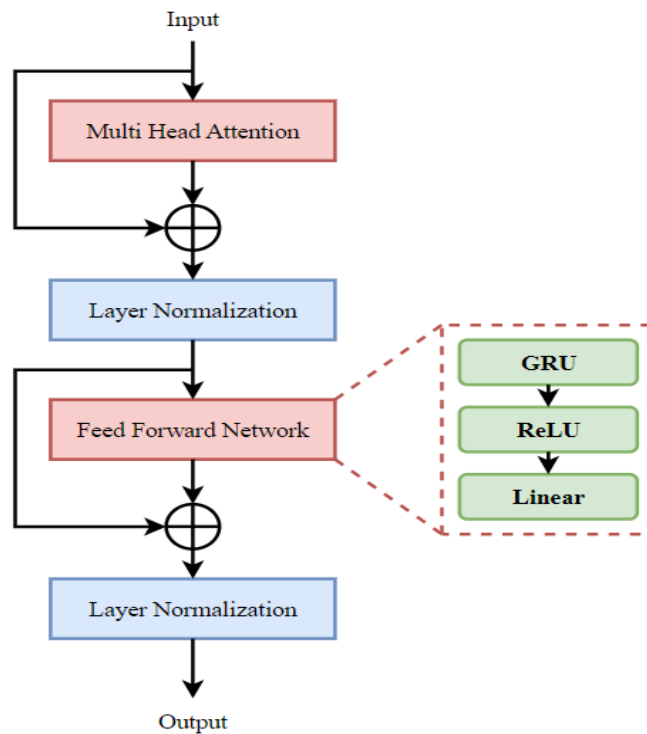
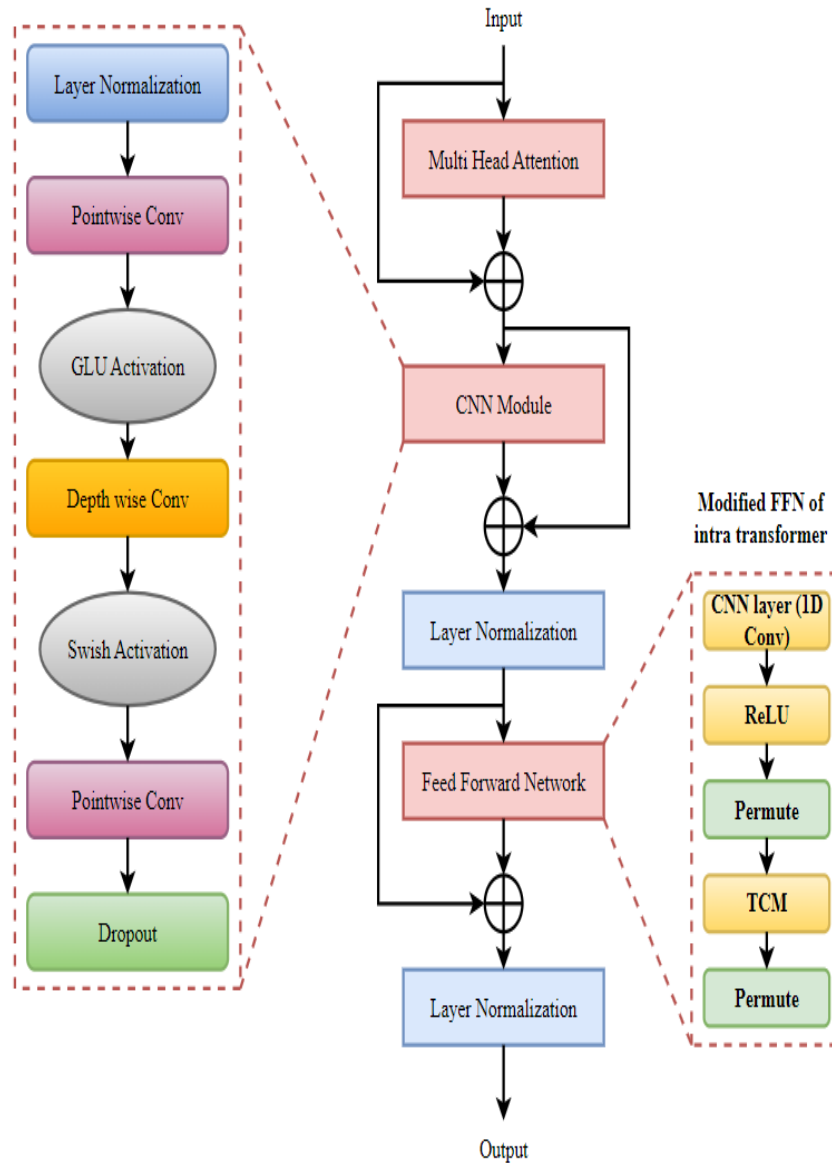


Fig. 3. The structure of feed forward network in baseline model

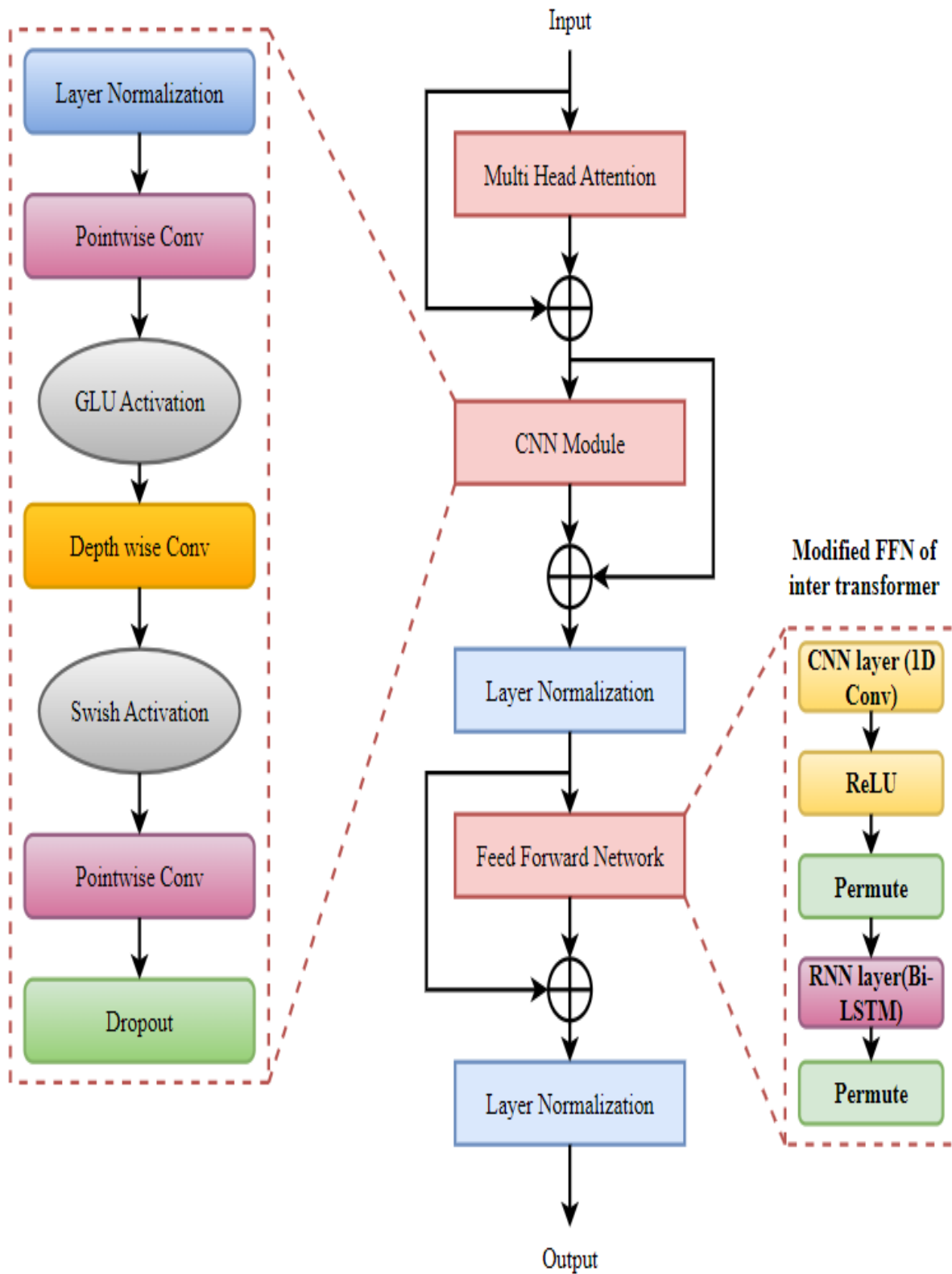
In the baseline model same structure of feed forward network is used in both intra and inter transformers of transformer block. In the proposed method unlike baseline, we introduced two different feed forward network layer structures in the intra and inter transformers of transformer block to well model the local and global features. In addition to this we have also added a convolutional module to the baseline intra and inter transformers of transformer block to further filter the local (short-term) noises present at the output of encoder. Due to the powerful feature self-learning ability of the CNN model, it will increase the local feature extraction ability of the transformer model. The proposed model enhances the local information modelling ability of the traditional transformer model by adding a convolution layer. The structure of proposed intra and inter transformers along with convolutional module are shown in below Fig. 4(a) and 4(b). The

convolutional module consists of Depth and Point wise convolutions instead of normal convolution to increase the speed of operation of model and both of them have less parameters compared to normal convolution.

2) *Proposed intra and inter transformer:* The key distinction between intra and inter transformer is in its structure of Feed Forward Network (FFN). A CNN layer (1D-Conv) and temporal convolution module (TCM) are used in intra transformer to learn long-term dependency of speech from the encoder output. The layer details of TCM are shown in below Fig. 5. A CNN layer (1D-Conv) and Bi-directional long short-term memory (Bi-LSTM) [34] are used in inter transformer to extract long-distance global features.



(a)



(b)

Fig. 4. The structure of inter transformer in transformer block

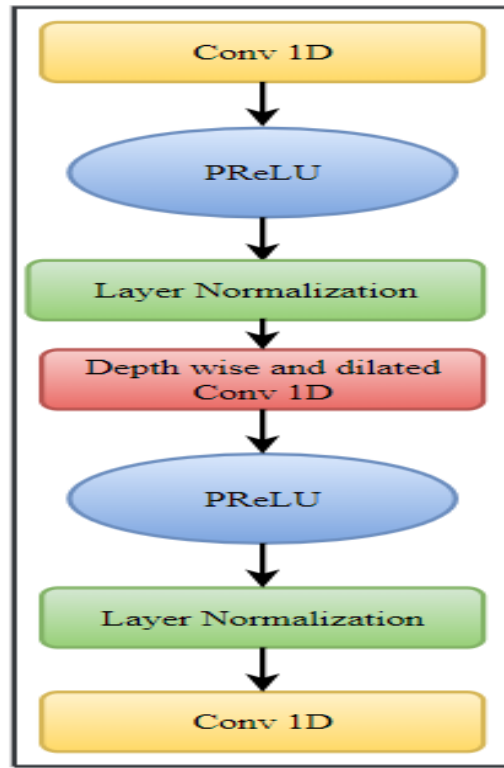


Fig. 5. The Layers of TCM

In general, Multi-Head Attention (MHA) works as follows:

In the MHA block, the input (Y) is first mapped  $h$ -times with different, learnable linear transformations to obtain queries, keys, and values representations, respectively.

$$Q_i = YW_i^Q, K_i = YW_i^K, V_i = YW_i^V \quad (2)$$

In the above Eq. 1  $Q_i, K_i, V_i \in R^{l \times d/h}$  are mapped query, key and value.  $W_i^Q, W_i^K, W_i^V \in R^{d \times d/h}$  represents  $i^{th}$  linear projection matrix for query, key and value.

The query will be dot produced with all keys and a constant is divided by the dot product. After that a softmax is applied to the values to obtain weights. According to Eq.2, each head's attention is a dot product of its weight and value.

$$Head_i = Attention(Q_i, K_i, V_i) = softmax\left(\frac{Q_i K_i^T}{\sqrt{d}}\right) V_i, i = 1, 2, \dots, h \quad (3)$$

The number of parallel attention heads are represented by  $h$ . In our proposed work  $h$  is considered as 4. The final output is obtained by concatenating the attentions of all heads and linearly projecting them again.

$$MultiHead(Q, K, V) = concat(Head_1, \dots, Head_h)W^O \quad (4)$$

In the above Eq. 3  $W_i^O \in R^{d \times d}$  is linear projection matrix.

Next, it follows residual connection as shown in Eq. (4).

$$out = Y + MultiHead(Q, K, V) \quad (5)$$

In the above Eq. 4  $Y \in R^{l \times d}$  represents input. Where  $l$  is sequence length and  $d$  is dimension.

To get the final output of intra and inter transformers of transformer block, the output of multi-head attention block is then processed by convolutional module and corresponding feed-forward networks, as well as residual connections and layer normalization [18].

The output of convolutional module is given as

$$\begin{aligned} Conv\ Module\ (out) \\ = Dropout(PointwiseConv(Swish(DepthWiseConv \\ (GLU(PointwiseConv \\ (LayerNormalization(out))))))) \quad (6) \\ CM_{out} = \\ LayerNormalization(out + (Conv\ Module\ (out))) \quad (7) \end{aligned}$$

The final output of intra transformer is given as

$$FFN_{intra}(CM_{out}) = TCN(ReLU(Conv1D(CM_{out}))) \quad (8)$$

$$Finaloutput_{intra} = LayerNormalization(CM_{out} + FFN_{intra}(CM_{out})) \quad (9)$$

The final output of inter transformer is given as:

$$FFN_{inter}(CM_{out}) = \text{Bi-LSTM}(\text{ReLU}(\text{Conv1D}(CM_{out}))) \quad (10)$$

$$\text{Finaloutput}_{inter} = \text{LayerNormalization}(CM_{out} + FFN_{inter}(CM_{out})) \quad (11)$$

3) *Proposed two-stage transformer block*: In the proposed 2-stage transformer block there are two transformers called intra and inter transformers as shown in Fig. 2, which are used for extracting local and global context information respectively. The local transformer with the input as a 3-D tensor with dimensions as [C, N, F] is first applied to each chunk and local information is processed in parallel on the last dimension F of the input tensor. For learning global dependency, the global transformer uses the information of output from the local transformer which is implemented on tensor dimension N. Moreover, each transformer undergoes a group normalization and residual connections are also used.

#### D. Masking Module

A masking network obtains denoising masks by utilizing transformer module's output features. First to match the output of the encoder, transformer module output is doubled along the channel dimension using PReLU and convolution. Afterwards, it undergoes two-way 2D convolution and nonlinearity, with the outputs multiplying together to form the input of two-dimensional convolution and PReLU. In order to obtain the final masked encoder feature, the mask and the encoder's output are multiplied element-wise.

#### E. Decoder

At this stage decoding the encoder feature into enhanced speech features is accomplished via dilated dense blocks and sub-pixel convolutions [35]. Using 2D Conv with filter size (1, 1), the enhanced speech feature's channel dimension is recovered into one and the enhanced waveform is produced by overlap-add.

#### F. Loss Function

The loss in the T-F domain can direct the model to acquire extra information, resulting in improved intelligibility as well as quality [19]. To train the model, we combine two losses. First, a waveform is created using the overlap-and-add approach utilizing the improved frames. Using the mean squared error among the enhanced and clean utterances, an utterance level loss is computed in the temporal domain.

It is said that the time-domain loss is:

$$L_t(y, \hat{y}) = \frac{1}{M} \sum_{n=0}^{M-1} (y_i[n] - \hat{y}_i[n])^2 \quad (12)$$

where M is the utterance length,  $y[n]$  is the nth sample of the clean utterance and  $\hat{y}[n]$  is nth sample of enhanced utterance.

The loss in frequency-domain is given as:

$$L_f(y, \hat{y}) = \frac{1}{T \cdot F} \sum_{t=1}^T \sum_{f=1}^F (|Y(t, f)_r| + |Y(t, f)_i|) - (|\hat{Y}(t, f)_r| + |\hat{Y}(t, f)_i|) \quad (13)$$

where  $Y(t, f)$  and  $\hat{Y}(t, f)$  denotes T-F units of STFT's of Y and  $\hat{Y}$  respectively. Where T denotes total frames number

and F denotes frequency bins count. The real and imaginary components of a complex variable Y is denoted by  $Y_r$  and  $Y_i$ , respectively.

Finally, the T-F domain loss is given by

$$L(y, \hat{y}) = \alpha * L_t(y, \hat{y}) + (1 - \alpha) * L_f(y, \hat{y}) \quad (14)$$

The hyper-parameter  $\alpha$  is set to 0.2.

## IV. EXPERIMENTS

### A. Datasets

We conducted the experiments using a 2 data sets one is public dataset published by Valentini et al. in [36]. This database contains 30 utterers from the Voice Bank corpus [37], among them 28 utterers are utilized to train the model and 2 are utilized for testing. There are 11,572 pairs of clean-noisy speeches in the training set. The noisy environments comprise 10 types of noises. Among them 8 noises are from the DEMAND dataset [38] and 2 are synthetic sounds at SNRs of 0, 5, 10 and 15 dB. The test set contains 824 mixtures using 5 noises from [38] which are not present in training set at SNRs of 17.5 dB, 12.5 dB, 7.5 dB and 2.5 dB.

Additionally, we assess the performance of our model in unseen environments. To test the generalization capability of proposed model we considered large Librispeech dataset [39]. To train the model we have chosen clean speech with duration of 50h from train-clean-100. A randomly selected set of noises from DEMAND dataset [38] and music samples from MUSAN [40] was used in the training set, with SNR ranges from -10dB to 10 dB. For each of the six DEMAND categories, we used two out of three types of noise. The test set of 500 samples are randomly taken from test-clean and the babble noise from NOISEX-92 dataset [41] and the noises river and restaurant are taken from DEMAND which are different from training set. The SNR are -5dB, 0dB, 5dB, 10dB, and 15dB. The sampling rate of all utterances is 16KHz.

### B. Experimental Setup

Each speech is resampled to a 16 kHz frequency. A rectangle window with a 32 ms (512 samples) size and overlap length of 16 ms is used to extract the frames. If an utterance lasts more than 4 seconds, we chunk a random 4 second chunk from it during each training session. To make the smaller utterances in the batch the same size as the largest utterance, zero padding is used. we train the model over the course of 100 epochs and optimization employs the Adam optimizer. Gradient clipping is used with a maximum L2-norm of 5 to prevent gradient explosion. During the training phase, dynamic strategies are used to determine learning rate [13]. In this experiment, learning rates are linearly increased during training and then reduced by 0.98 every two epochs.

### C. Evaluation Metrics

In this study, speech quality is measured by using STOI [42], whose values normally fall between 0 and 1, and PESQ [43], whose ranges are -0.5 to 4.5. Speech quality improves with a higher PESQ value. Speech intelligibility increases with increased STOI.

#### D. Results and Analysis

1) *Comparison with existing works on Valentini et al. dataset [36]*: The experiment was carried out with two different datasets. The proposed convolutional transformer model was compared with various existing models such as SEGAN [8], CGAN [44], Wave-U-Net [7], MMSE-GAN [2], Metric GAN [3], DCUnet-16 [6], DEMUCS-small [11], SE-Transformer [27], TSTNN [28]. All the existing models used

for comparison are trained with similar dataset used to train our model. All the models are reproduced for comparison. As a comparison to the number of existing models stated in the original papers, we calculated the metric scores of our model. The proposed model was assessed by using the STOI and PESQ. The PESQ and STOI values for Valentini et al. dataset [36] are given in Table I.

TABLE I. ASSESSMENT OF PESQ AND STOI OF PROPOSED MODEL WITH PREVIOUSLY PUBLISHED BASELINE SCORES USING THE VALENTINI ET AL. DATASET

MODEL	PESQ	STOI (%)
Noisy	1.97	91
SEGAN [8]-2017	2.16	93
CGAN [44]-2018	2.34	93
Wave-U-Net [7]-2018	2.40	-
MMSE-GAN [2]-2018	2.53	93
Metric GAN [3]-2019	2.86	-
DCUnet-16 [6]-2019	2.93	-
DEMUCS-small [11]-2020	2.93	95
SE-Transformer [27]-2022	2.62	93
TSTNN [28]-2021	2.96	95
Convolutional TSTNN (proposed)	3.12	96

The above Table I gives the PESQ and STOI scores for proposed and various existing models such as SEGAN [8], CGAN [44], Wave-U-Net [7], MMSE-GAN [2], Metric GAN [3], DCUnet-16 [6], DEMUCS-small [11], SE-Transformer [27], and TSTNN [28].

The PESQ score for SEGAN [8] is 2.16 and for the proposed is 3.12. The STOI score for SEGAN [8] is 93% and for the proposed is 96%. The SEGAN is an end-to-end SE model where only strided convolutions are used in the generator and discriminator. The PESQ score for CGAN [44] is 2.34 and for the proposed is 3.12. The STOI score for CGAN [44] is 93% and for the proposed is 96%. The CGAN is a CNN based GAN operates in T-F domain. In both the GAN models only, ordinary convolutional layers are used. The PESQ score for DEMUCS-small [11] is 2.93 and for the proposed is 3.12. The STOI score for DEMUCS-small [11] is 95% and for the proposed is 96%. The DEMUCS model is a time domain U-Net model with ordinary convolutions and LSTM bottle neck. The CNN alone cannot well model the long-range dependencies of speech signal. In the proposed model to further enhance the performance of the model in addition to ordinary convolution a dilated dense block is used

in both encoder and decoder. At various resolutions, the dilated convolutions aid with context aggregation and the dense connectivity provides feature map with more precise target information by passing through multiple layers. A transformer block is used for extracting the local and global features from a noisy speech signal. The PESQ score for SE-Transformer [27] is 2.62 and for the proposed is 3.12. The STOI score for SE-Transformer [27] is 93% and for the proposed is 96%. The SE-transformer uses only LSTMs, multi-head attention and 1D convolution for SE. The PESQ score for TSTNN [28] is 2.96 and for the proposed is 3.12. The STOI score for TSTNN [28] is 95% and for the proposed is 96%.

2) *Comparison with existing works on librispeech dataset [39]*: To test the generalization ability of our proposed model we conducted experiments with large Librispeech dataset. The average PESQ and STOI scores for three types of noises at SNRs of -5dB, 0dB, 5dB, 10dB, and 15dB are presented in Table II. The proposed convolutional transformer model was compared with DNS-baseline [45], DEMUCS [11] and TSTNN [28].



TABLE II. EVALUATION OF PESQ AND STOI OF OUR MODEL AND EXISTING MODELS IN THE EXISTENCE OF BABBLE NOISE ON LIBRISPEECH DATASET

Model	Noisy		DNS [45]		DEMUCS [11]		TSTNN [28]		Proposed	
	PESQ	STOI	PESQ	STOI	PESQ	STOI	PESQ	STOI	PESQ	STOI
<b>-5dB</b>	1.11	0.66	1.23	0.70	1.50	0.82	1.58	0.83	1.69	0.84
<b>0dB</b>	1.19	0.77	1.48	0.82	1.92	0.90	2.10	0.91	2.32	0.92
<b>5dB</b>	1.40	0.86	1.86	0.90	2.36	0.94	2.49	0.94	2.68	0.94
<b>10dB</b>	1.75	0.92	2.29	0.94	2.76	0.96	2.88	0.97	3.04	0.96
<b>15dB</b>	2.25	0.96	2.71	0.96	3.07	0.98	3.19	0.98	3.58	0.98

From Table II, in the existence of babble noise at -5dB input SNR, the PESQ value for the proposed method is 1.69 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 1.23, 1.50 and 1.58. The STOI score of the proposed method is 0.84 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 0.70, 0.82 and 0.83.

From Table II, in the existence of babble noise at 5dB input SNR, the PESQ value for the proposed method is 2.68 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 1.86, 2.36 and 2.49. The STOI score of the proposed method is 0.94 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 0.90, 0.94 and 0.94.

From Table II, in the existence of babble noise at 15dB input SNR, the PESQ value for the proposed method is 3.58 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 2.71, 3.07 and 3.19. The STOI score of the proposed method is 0.98 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 0.96, 0.98 and 0.98.

From Table II in the existence of babble noise at all input SNR conditions it is noted the proposed model performs better than that of DNS baseline [45], DEMUCS [11] and TSTNN [28] models. The proposed convolutional transformer delivers improved performance because the local and global features are well modelled by CNN and RNN layers. Moreover, a CNN module is added to deal with short-term noise.

TABLE III. EVALUATION OF PESQ AND STOI OF PROPOSED MODEL AND EXISTING MODELS IN THE EXISTENCE OF RIVER NOISE ON LIBRISPEECH DATASET

Model	Noisy		DNS [45]		DEMUCS [11]		TSTNN [28]		Proposed	
	PESQ	STOI	PESQ	STOI	PESQ	STOI	PESQ	STOI	PESQ	STOI
<b>-5dB</b>	1.23	0.80	1.81	0.86	2.05	0.90	2.12	0.91	2.28	0.92
<b>0dB</b>	1.45	0.87	2.17	0.91	2.37	0.94	2.46	0.94	2.67	0.95
<b>5dB</b>	1.78	0.92	2.55	0.95	2.66	0.96	2.84	0.96	2.96	0.96
<b>10dB</b>	2.23	0.96	2.92	0.96	3.98	0.97	3.12	0.97	3.42	0.97
<b>15dB</b>	2.78	0.98	3.23	0.98	3.33	0.98	3.41	0.98	3.64	0.98

From Table III, in the existence of river noise at 0dB input SNR, the PESQ value for the proposed method is 2.67 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 2.17, 2.37 and 2.46. The STOI score of the proposed method is 0.95 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 0.91, 0.94 and 0.94.

From Table III, in the existence of river noise at 5dB input SNR, the PESQ value for the proposed method is 2.96 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 2.55, 2.66 and 2.84. The STOI score of the proposed method is 0.96 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 0.95, 0.96 and 0.96.

From Table III, in the existence of river noise at 10dB input SNR, the PESQ value for the proposed method is 3.42 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 2.92, 3.98 and 3.12. The STOI score of the proposed method is 0.97 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 0.96, 0.96 and 0.97.

From Table III in the existence of river noise at all input SNR conditions it is noted the proposed model performs better than that of DNS baseline [45], DEMUCS [11] and TSTNN [28] models. The proposed convolutional transformer delivers improved performance because the local and global features are well modelled by CNN and RNN layers. Moreover, a CNN module is added to deal with short-term noise.

TABLE IV. EVALUATION OF PESQ AND STOI OF PROPOSED MODEL AND EXISTING MODELS IN THE EXISTENCE OF RESTAURANT NOISE ON LIBRISPEECH DATASET

Model	Noisy		DNS [45]		DEMUCS [11]		TSTNN [28]		Proposed	
	PESQ	STOI	PESQ	STOI	PESQ	STOI	PESQ	STOI	PESQ	STOI
-5dB	1.09	0.63	1.21	0.68	1.38	0.78	1.49	0.79	1.62	0.81
0dB	1.15	0.75	1.43	0.80	1.68	0.88	1.85	0.91	1.98	0.92
5dB	1.31	0.84	1.76	0.88	1.99	0.93	2.29	0.94	2.41	0.94
10dB	1.62	0.91	2.16	0.93	2.33	0.95	2.67	0.95	2.87	0.96
15dB	2.07	0.95	2.59	0.96	2.70	0.97	2.94	0.97	3.18	0.97

From Table IV, in the existence of restaurant noise at 0dB input SNR, the PESQ value for the proposed method is 1.98 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 1.43, 1.68 and 1.85. The STOI score of the proposed method is 0.92 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 0.80, 0.88 and 0.91.

From Table IV, in the existence of restaurant noise at 10dB input SNR, the PESQ value for the proposed method is 2.67 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 2.16, 2.33 and 2.67. The STOI score of the proposed method is 0.96 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 0.93, 0.95 and 0.95.

From Table IV, in the existence of restaurant noise at 15dB input SNR, the PESQ value for the proposed method is 3.18 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 2.59, 2.70 and 2.94. The STOI score of the proposed method is 0.97 and for existing DNS baseline [45], DEMUCS [11] and TSTNN [28] are 0.96, 0.97 and 0.97.

From Table IV in the existence of restaurant noise at all input SNR conditions it is noted the proposed model performs better than that of DNS baseline [45], DEMUCS [11] and TSTNN [28] models. The proposed convolutional transformer delivers improved performance because the local and global features are well modelled by CNN and RNN layers. Moreover, a CNN module is added to deal with short-term noise.

From Table II, Table III, and Table IV, it is noted the transformer-based models performs better than that of DNS baseline [45] and DEMUCS [11]. The proposed convolutional transformer delivers improved performance in terms of enhanced signal quality than all existing methods for all types of noises at all SNR conditions.

## V. DISCUSSION

The results from experimental studies show that the proposed convolutional transformer consistently improves speech quality as well as intelligibility. The performance of proposed transformer is better than the existing models such as SEGAN, CGAN, Wave-U-Net, MMSE-GAN, Metric GAN, DCUnet-16 and TSTNN. In the existing models CNN alone cannot well model the long-range dependencies of speech signal. In the proposed model to further enhance the performance of the model in addition to ordinary convolution

a dilated dense block is used in both encoder and decoder. At various resolutions, the dilated convolutions aid with context aggregation and the dense connectivity provides feature map with more precise target information by passing through multiple layers. A transformer block is used for extracting the local and global features from a noisy speech signal. In TSTNN model same structure of FFN is used in both intra and inter transformers. And also, they have not concentrated on the local (short-term) noises present at the output of encoder. To further enhance the performance in the proposed model we used TCM in the intra transformer and Bi-LSTM in the inter transformer as CNN's are good at extracting local features and the RNN's are good at extracting global features from the noisy speech signal. In the proposed model the transformer considers global information as well as parallel computing, resulting in a reduction of long-term noise. In contrast to RNNs and CNNs, the proposed transformer model uses parallel processing and the long-term dependencies can be modelled more efficiently with multi-head attention (MHA) by using sequence similarity. Moreover, a CNN module is added to the transformer so that short-term noise can be reduced more effectively, based on the ability of CNN to extract local information. We observe that the performance drop is more without CNN module that means on the basis of the transformer model CNN module is able to capture locally relevant context information based on global contextual information.

## VI. CONCLUSION

In this paper we proposed a two-stage convolutional transformer for speech enhancement in time domain. The transformer considers global information as well as parallel computing, resulting in a reduction of long-term noise. In the proposed work unlike two-stage transformer neural network (TSTNN) different transformer structures for intra and inter transformers are used for extracting the local as well as global features of noisy speech. Moreover, a CNN module is added to the transformer so that short-term noise can be reduced more effectively, based on the ability of CNN to extract local information. The experimental findings demonstrate that the proposed model outperformed the other existing models on both the datasets in terms of STOI (short-time objective intelligibility), and PESQ (perceptual evaluation of the speech quality). In future, we would like to analyze the performance of the transformer in T-F domain by adding new layers like Time Frequency attention (TFA) and also apply it in speech separation and multi-channel speech enhancement.

REFERENCES

- [1] S.-W. Fu, T.-y. Hu, Y. Tsao, and X. Lu, "Complex spectrogram enhancement by convolutional neural network with multi-metrics learning," in 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2017, pp. 1–6
- [2] M. H. Soni, N. Shah, and H. A. Patil, "Time-Frequency Masking-Based Speech Enhancement Using Generative Adversarial Network," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 5039–5043.
- [3] S.-W. Fu et al., "Metricgan: Generative adversarial networks based black-box metric scores optimization for speech enhancement," in ICML, 2019.
- [4] N. Shah, H. A. Patil, and M. H. Soni, "Time-Frequency Mask-based Speech Enhancement using Convolutional Generative Adversarial Network," in Proceedings, APSIPA Annual Summit and Conference, 2018, vol. 2018, pp. 12–15.
- [5] N. Takahashi, P. Agrawal, N. Goswami, and Y. Mitsufuji, "PhaseNet: Discretized Phase Modeling with Deep Neural Networks for Audio Source Separation," in Proc. Interspeech 2018, 2018, pp. 2713–2717.
- [6] H.-S. Choi, J.-H. Kim, J. Huh, A. Kim, J.-W. Ha, and K. Lee, "Phase-aware Speech Enhancement with Deep Complex UNet," Mar. 2019.
- [7] C. Macartney and T. Weyde, "Improved speech enhancement with the wave-u-net," arXiv preprint arXiv:1811.11307, 2018.
- [8] Santiago Pascual, Antonio Bonafonte, and Joan Serra, "Segan: Speech enhancement generative adversarial network," arXiv preprint arXiv:1703.09452, 2017.
- [9] D. Rethage, J. Pons, and X. Serra, "A Wavenet for Speech Denoising," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 5069–5073.
- [10] A. Pandey and D. Wang, "TCNN: Temporal convolutional neural network for real-time speech enhancement in the time domain," in ICASSP, 2019, pp. 6875–6879.
- [11] Defossez A, Synnaeve G, Adi Y, "Real Time Speech Enhancement in the Waveform Domain". arXiv preprint arXiv:2006.12847, 2020.
- [12] Fu S, W Tsao, Y Lu, X & Kawai H., "Raw waveform-based speech enhancement by fully convolutional networks," In Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, pp. 006-012, IEEE, 2017.
- [13] H Erdogan, J R Hershey, S Watanabe, and J Le Roux., "Phase-sensitive and recognition-boosted speech separation using deep recurrent neural networks," In ICASSP, pp. 708–712, 2015.
- [14] Y Wang, A Narayanan, and D Wang., "On training targets for supervised speech separation," IEEE/ACM Transactions on Audio, Speech and Language Processing, vol. 22, no. 12, pp. 1849–1858, 2014.
- [15] Tan K, & Wang D "A convolutional recurrent neural network for real-time speech enhancement,". In Interspeech, Vol. 2018, pp. 3229-3233, 2018.
- [16] Yu F, Koltun V, "Multi-scale context aggregation by dilated convolutions." arXiv preprint arXiv:1511.07122, 2015.
- [17] K Tan, J Chen, and D L Wang., "Gated residual networks with dilated convolutions for monaural speech enhancement," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 27, pp. 189–198, 2019.
- [18] J Chen and D L Wang., "Long short-term memory for speaker generalization in supervised speech separation,". The Journal of the Acoustical Society of America, vol. 141, no. 6, pp. 4705–4714, 2017.
- [19] Pandey and D Wang., "A new framework for supervised speech enhancement in the time domain," In Proceedings of Interspeech, pp. 1136–1140, 2018.
- [20] Pandey A, & Wang D., "Densely connected neural network with dilated convolutions for real-time speech enhancement in the time domain," In International Conference on Acoustics Speech and Signal Processing, pp. 6629-6633, IEEE,2020.
- [21] Li Qinglong, Fei Gao, Haixin Guan, and Kaichi Ma., "Real-time monaural speech enhancement with short-time discrete cosine transform,". arXiv:2102.04629,2021.
- [22] K Zhen, M S Lee and M Kim., "A Dual-Stage Context Aggregation Method towards Efficient End-to-End Speech Enhancement," International Conference on Acoustics, Speech and Signal Processing, pp. 366-370,2020.
- [23] Lin Ju, Adriaan J. Van Wijngaarden, Melissa C. Smith, and Kuang-Ching Wang., "Speaker-Aware Speech Enhancement with Self-Attention," "In 29th European Signal Processing Conference (EUSIPCO), pp. 486-490. IEEE, 2021.
- [24] Pandey, A., & Wang, D. (2022). Self-attending RNN for speech enhancement to improve cross-corpus generalization. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30, 1374-1385.
- [25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in neural information processing systems, 2017, pp. 5998–6008.
- [26] J. Kim, M. El-Khamy, and J. Lee, "T-gsa: Transformer with gaussian-weighted self-attention for speech enhancement," in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 6649–6653.
- [27] W. Yu, J. Zhou, H. Wang, and L. Tao, "SETransformer: Speech enhancement transformer," Cogn. Comput., vol. 2021, pp. 1–7, Feb. 2021.
- [28] K. Wang, B. He, and W. Zhu, "TSTNN: Two-stage transformer based neural network for speech enhancement in the time domain," in Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP) Conf., 2021, pp. 7098–7102.
- [29] Y. Luo, Z. Chen, and T. Yoshioka, "Dual-path rnn: efficient long sequence modeling for time-domain single-channel speech separation," in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 46–50.
- [30] Zhao, J., Mao, X., & Chen, L., "Speech emotion recognition using deep 1D & 2D CNN LSTM networks," Biomedical signal processing and control, 47, 312-323.
- [31] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in IEEE International Conference on Computer Vision, 2015, pp. 1026–1034.
- [32] M. Sperber, J. Niehues, G. Neubig, S. Stuker, and A. Waibel, "Self-attentional acoustic models," Proc. Interspeech 2018, pp. 3723–3727, 2018.
- [33] J Chen, Q Mao, D. Liu "Dual-Path Transformer Network: Direct Context-Aware Modeling for End-to-End Monaural Speech Separation". arXiv preprint arXiv:2007.13975, 2020.
- [34] Li, X., Li, Y., Dong, Y., Xu, S., Zhang, Z., Wang, D. and Xiong, S., 2020. Bidirectional LSTM Network with Ordered Neurons for Speech Enhancement. In Interspeech (pp. 2702-2706).
- [35] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in IEEE conference on computer vision and pattern recognition, 2016, pp. 1874–1883.
- [36] C. Valentini-Botinhao et al., "Noisy speech database for training speech enhancement algorithms and tts models," 2017.
- [37] Veaux C, Yamagishi J, King S. "The voice bank corpus: Design, collection and data analysis of a large regional accent speech database." In 2013 International Conference Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation(O-COCOSDA/CASLRE). IEEE, 2013: 1-4.
- [38] Joachim Thiemann, Nobutaka Ito, and Emmanuel Vincent, "The diverse environments multi-channel acoustic noise database: A database of multichannel environmental noise recordings," The Journal of the Acoustical Society of America, vol. 133, no. 5, pp. 3591–3591, 2013.
- [39] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: an asr corpus based on public domain audio books," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015, pp. 5206–5210.
- [40] D. Snyder, G. Chen, and D. Povey, "Musan: A music, speech, and noise corpus," 2015.

- [41] A. Varga and H. J. Steeneken, "Assessment for automatic speech recognition: Ii. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, no. 3, pp. 247–251, 1993.
- [42] C H Taal, R C Hendriks, R Heusdens, and J Jensen., "An algorithm for intelligibility prediction of time– frequency weighted noisy speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, 2011.
- [43] W Rix, J G Beerends, M P Hollier, and A P Hekstra., "Perceptual evaluation of speech quality (PESQ) - a new method for speech quality assessment of telephone networks and codecs," In *ICASSP*, pp. 749–752, 2001.
- [44] Shah N, Patil A, Soni H. Time-frequency mask-based speech enhancement using convolutional generative adversarial network, *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. 2018; p. 1246–51.
- [45] C. K. Reddy, H. Dubey, V. Gopal, R. Cutler, S. Braun, H. Gamper, R. Aichner, and S. Srinivasan, "Icassp 2021 deep noise suppression challenge," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

# Assessing User Interest in Web API Recommendation using Deep Learning Probabilistic Matrix Factorization

T. Ramathulasi<sup>1</sup>, M. Rajasekhara Babu<sup>2\*</sup>

Research Scholar, School of Computer Science & Engineering, Vellore Institute of Technology, India<sup>1</sup>  
Professor, School of Computer Science & Engineering, Vellore Institute of Technology, India<sup>2</sup>

**Abstract**—Internet 2.0 Things connected to the Internet not only manage data supply through devices but also control the commands that flow through it. The communication technology created by the desired sensor is used by a new computing model so that the collected data appears in Web 2.0 for management. In addition to enhancing Sense efficiency through the simple IoT computing process, it is used in many cases for example video surveillance, and improved and intelligent manufacturing. Every fragment of the system is carefully continued and supervised in this process by software collection using a large number of recurs. An important process for this is to access web APIs from various public platforms in an efficient way. The use of different APIs by developers for the integration of different IoT devices and the deployment process required for this is unnecessary. Obtaining configured target APIs makes it easy to know where and how to get started with the workflow approach. Rapid industrial development can be achieved through this powerful API approach. But finding adequately powerful APIs from a large number of APIs has become a great challenge. However, due to the massive spike in the count of APIs, combining the two APIs has now become a major challenge. In this paper, for the time being, only the relationships between users and the API are considered. In this case, they had to face difficulties in extracting contextual value from their interpretation. So better accuracy could not be obtained due to this. The consequence of the user's time aspect on the cryptographic properties concerning the information collected from the API contextual description can be enhanced by the Deep Learning Probabilistic Matrix Factorization (DL-PMF) method, which improves the accuracy of the API recommendation in considering the cryptographic features of the user in the API recommendation. In this paper, we have used CNN (Convulsive Neural Network) for web elements such as APIs, and LSTM (Long-Term and Short-Term Memory) Network, which works with a diligent mechanism to find hidden features, to find hidden features that suit the tastes of the users. In conclusion, the combination of PMF (Probabilistic Matrix Factorization) evaluation of the recommended results was obtained as described above. The combination of DL-PMF method experimental results was found to be better than previous PMF, ConvMF, and other methods, thus improving the recommended accuracy.

**Keywords**—Implicit feature; API's recommendation; IoT; collaborative filtering; matrix factorization

## I. INTRODUCTION

In this new era due to the huge explosion in data registration as an online platform, consumers have to face various big problems in big data usage. Retrieving large data and other necessary information by communicators using [1]

technology in a very short period has not been very successful. In addition to providing services tailored to the personal preferences of different users, this method also answers the major problem of data overload. This order, which was recommended by a mature process developed by a highly researched cooperative filtering method, became highly publicized. This has been referred to as neighborhood and model-based practices [2]. Better results than neighborhood-based recommended policy outcomes can be obtained by recommending using PMF and SVD methods with model-based approaches [3]. The probability factor established by the model-based approaches PMF and SVD plays a major role in gradually improving the effectiveness of the recommended process. The PMF approach turned out to be effective in uncovering the hidden features of consumers by rating matrix but has failed to make effective and efficient utilization of this informational interpretation in terms of useful and vital customer information and services.

In this case, the recommendation system was applied through approaches undertaken by several researchers with the knowledge gained from the in-depth study. As a result, the recommended system has been improved by eliminating the flaws in the traditional recommendation method and utilizing information efficiently through complex approaches. There are still some errors that need to be corrected. Two CNNs (Conventional Neural Networks) were created in parallel to capture the underlying elements of Web API description information and the factor evaluation mechanism from the customer evaluation description [4]. The impact of consumer preferences on the recommendation system could not be ascertained due to the time factor not taking into account the consumer preferences in this model. For better recommendation, we have combined both the denoising autoencoder and the CDL model, which can effectively obtain encryption components through auxiliary information, and achieve better results [5]. The hierarchical Bayesian model approach is proposed later in this series.

The methods described by the above research have not been able to achieve better results in extracting relevant information from the content description. CNN and PMF have been used interchangeably, as described in the research papers [6] and [7] to identify latent properties as the basis of object content. But this does not allow us to find out the latent features of the user using the user information. The model [8], developed by combining CNN and latent factors, was successful in capturing users' object attributes but failed to achieve better-optimized hidden attributes. RNN (Repeated

\*Corresponding Author.

Neural Network) assists in the detection of consumer data irregularities [9] and the output of received news messages for recommendation by taking what is collected and received as input for what was previously considered the Browsing History Helpline. Dimensionally we get output messages that have the same dimension size as the input in the feature learning process. This requires the use of optimized dimensionality. In addition to extracting latent features from the user information description, the time it takes to use them in the process is also calculated using the RNN-based model [10] proposed by the collaborative filtering algorithm. Failed to fully extract from the elements the comprehensive latent features based on the model described in this sequence.

It has now become a challenge to capture the latent features according to their real, precise needs according to the information in the description of the consumer items. Integrating the PMF model with Deep Learning has been proposed for its solution. The text descriptive for this system is given as input for recommendation and the whole process is done in three steps. The latent features were learned from the description of the elements in the first stage, the latent features of the users in the second stage, and finally, the total latent features were trained by merging the two latent features into the PMF. To this end, in the aforementioned paper, we have attached a technique with particular care to capture and better understand the exact latent nature of CNN from element description [11, 12]. Care is taken to remove useless information in the service description as well as to learn the latent features according to the tastes of the customers [13]. In this paper, LSTM has been adapted LSTM to find the latent features appropriate to the needs of real and accurate users [14, 15]. The impact of consumption time on consumer pre-history discussions and consumer tastes has been taken into account in the process of learning these hidden features. Finally, it was made possible by combining users' hidden features with PMF for content-based assessment of user preference using both elements and user description. The combination of ConvMF and PMF has been verified by proposing Deep Learning-based PMF for greatly improved accuracy in the recommendation and enhanced accuracy capability by experimenting on a data set that has been crawled from a website called Programmableweb.

## II. RELATED WORK

The section includes, the most widely used CF algorithms in service recommendation [16, 17, 18] are currently reviewed by Mac as algorithms relevant to service and API recommendations. In this paper, Mache proposes a hybrid CF algorithm to overcome the shortcomings of CF algorithms [17] following current traditional methods. For this purpose, the information of the web APIs was collected by the first collecting mechanism and a new method of weight calculation was used. This algorithm then improved the accuracy of the recommendation by combining the user and item-based CF model with different weights. The web API information description was collected by a mechanism proposed in [16]. Different weights were added to the user and item-based CF model through the new-weight calculation method.

The revised proposal has sparked renewed interest in APIs in recent years as a means of addressing the guideline's flaws. The algorithms in this sector are separated into API and Mashup suggestions. The authors of [19] advised mashups for consumers based on customer preferences and service social networks for Mashup Recommendation. Historical usage data from the mashup is used to determine user interest. For the mashup service, a social network of web services was developed using a description of the information included in the API and associated tags. The network has uncovered and integrated user interest in web services. As noted in [16], potential affiliates have been discovered as a set of API mashups produced by authors from among APIs. It also focuses on an excavation in the context of Mashup's proposal for a shared API invoice model, as implemented by the excavated association recommended model. The potential model here improves the streamlined latent correlation.

The framework based on the Knowledge Graph was submitted by the authors of [20] on the recommendation of the APIs. It first recognizes the invite relationship between APIs and also identifies key information related to categories and labels. This information is used to form the relationship component between entities. The input is part of the recommendation in the knowledge graph. The authors later [21] proposed the recommendation of the framework obtained by applying random walks on the Knowledge Graph. The information on the most useful APIs was first compiled by the authors using the knowledge graph and the nodes were filtered based on the semantic information contained. The relationship between the candidate and the nominee was re-evaluated by the authors using random walks. Different networks are used to resolve different relationships represented by multiple links and multiple types of objects. A diverse information network is built by the authors by combining the API information with the information contained in [22] about the mashup and its related features, and group preference can be obtained for the mashup by following a clustering approach to the information in the network. Fixed issues with authors [23] currently having problems getting permission to search for API representative services by search engines. In addition, the recovery process was proposed by the new authors with the help of a collaborative network of services and the visualization technique provided for easier browsing.

Recommendations for APIs are made by traditional CF algorithms. The CF algorithm was proposed by the author in a paper [24] due to an item-based API recommendation. Item similarity can be achieved by making the customer group based on customers who have similar preferences. Furthermore, the CF algorithm applies a specific item based on each user group. It has been confirmed that better performance than neighbor-based CF can be achieved with MF model-based CF, with the authors attempting to assess user preferences when presenting invitations in [25] and terms of APIs and mashups. This assessment was made possible by the authors incorporating user preferences into the MF model to create a new recommendation model.

Even though recent procedures have a high level of suggestion accuracy, the content of the data has not been accomplished. Because most consumers use different objects at

various times throughout the day, Chen et al. [26] argued that contextual factors are crucial in the brilliant object suggestion of the Internet of Things. The researchers projected a time-aware, bright object recommendation model integrating user interest across times and creative object resemblance. Duan et al. [27] suggested JointRec, a deep learning-based joint video streaming recommendation system for IoT, to deliver precise video ways for a better marginal of users. Item descriptor information in IoT is generally diverse and multimodal, according to Huang et al. [28]. The paper presents a new heterogeneous representation learning-based strategy to improve predictive performance in IoT. In [29], Hu et al. suggested automotive ad-hoc network representation training for IoT recommendations since trajectory data better knowledge of human movement patterns. In contrast to these approaches, we investigate the potential linkages between users and API attributes in this research. To uncover implicit information, we use collaborative learning. Finally, we propose multiple models for IIoT devices and API recommendations grounded on the association between users and APIs. A clustering technique called Word-Embedded Clustering (AUG-LDA) is suggested in [30] for improved outcomes with a significant influence on word vector quality. This method can offer more efficiency and accuracy in mashup service innovation. DMM-Gibbs (Dirichlet Multinomial Mixer with Gibbs) predicts better clustering performance by reducing the number of features represented [31].

### III. PRIOR KNOWLEDGE

In this segment, we critically summarize the most popular collaborative filtering technique called “Matrix Factorization (MF)” and the “Convolutional Neural Network (CNN)”.

The MF method operates in a process-sharing cryptocurrency that calculates the strengths of the user-object relationship with the rating given by the consumer on the object, intending to collect hidden patterns of the object [3]. Let's assume we have  $N$  consumers and  $S$  web services and a ranking index for user service is  $R \in \mathbb{R}^{N \times S}$ . In MF,  $k$ -dimensional structures,  $u_i \in \mathbb{R}^k$  and  $s_j \in \mathbb{R}^k$  are represented as the latent structures of consumer  $i$  and service  $j$ . A user  $i$  and service  $j$ 's ranking  $r_{ij}$  is approximated by the inner-product of the related latent user  $i$  and service  $j$  models (i.e.  $r_{ij} \approx \hat{r}_{ij} = u_i^T s_j$ ). One possible simple way to better minimize the damage function  $\mathcal{L}$  value is to train Latin models. It has a policy of having total-squared-error terms between ratings and actual ratings, and  $L_2$  regularized terms. To better avoid the major problem of over-fitting, try to improve it as shown below:

$$\mathcal{L} = \sum_i^N \sum_j^S I_{ij} (s_{ij} - u_i^T s_j)^2 + \lambda_u \sum_i^N \|u_i\|^2 + \lambda_v \sum_j^S \|s_j\|^2 \quad (1)$$

Where an indicator function  $I_{ij}$  exists such that, if the user is given a rating it is 1 otherwise 0.

A "Convolutional Neural Network" (CNN) is called feed-forward and is a type of neural network that varies with the following components: 1) To construct local attributes, the convolution layer is used; 2) The pooling layer is used to represent only the data as an abbreviated representation, called the sub-template. Via activation functions, all of the

representatives with the highest score from the previous layer send local features to the next layer.

Although CNN was originally used to develop computer-vision-based applications, its main idea was to retrieve information and, in particular, to return the NLP search query [32], to model sentences [33], and additionally perform NLP tasks in traditional methods [34]. Here NLP has led to significant changes in the structure of CNN for the betterment of work and better results in the performance of the NLP's diverse work.

However, in reality, CNN is not yet well used in some areas, especially in the recommendation system. To the best of our knowledge, van den Ord et al. CNN [35] was first applied to music recommendation. There they analyzed the songs via CNN from a sound analysis perspective. And the desired pattern for calculating ratings is based on the latent pattern of the song obtained by sound CNN. The CNN model has not shown enough improvement in processing the documents and retrieving the sound signal. The reason for this is that the quality of the documents, the sound signals, and the features around them differ inherently with some variation. At a given time i.e. a signal with a short time difference is equal to the surrounding signals. And the word in a document at a certain point differs greatly from the meanings of the surrounding words. So, in the end, there is a great need for CNN to have such different variations and different structures to influence the quality of features locally. Also, the cooperative information of the model is not fully available. In particular, service latent features are determined by the analysis results obtained by CNN to convey the audio signal rather than the collaborative information. Therefore, "weighted matrix factorization (WMF)" [36] alone cannot improve the performance of the entire recommendation system, and it is one of the traditional collaborative filter MF-based methods that deals only with datasets with latent features.

### IV. METHODOLOGY OF DL-PMF MODEL

Matrix factors can be considered models by PMF as a potential solution. User preference in service selection is explained by the use of a potential combination of low-dimensional feature vectors, and potential properties are mapped by API and customer element information based on the potential angle to the low-dimensional mapping space. The process of identifying matrix factors as a potential solution was used as a model by the PMF. To assist in the dimensional mapping of a small area, it has used a low dimensional feature vector to highlight important information potential features in the API and user information. Specifically, this simple combination makes it possible to take customer preference into account in the process description recommended by the API. In this paper, the main purpose is to get better results using the proposed DL-PMF model. For this, the service / APIs and customer information through the continuous iteration process are used to find hidden vectors with the help of the PMF method. The performance of this model can be seen in a step-by-step sequence as shown in Table I below and Fig. 1 below.

TABLE I. THE ARRANGEMENT OF STEPS IN THE DEEP-LEARNING BASED PMF MODEL

Algorithm	The arrangement of steps in the Deep-Learning based PMF model
Step 1	Incidentally, both U and $W_S$ variables are generated and configured as rating matrix R.
Step 2	Attention-CNN updates S depending on R and U.
Step 3	S is trained using the PMF method while U is trained using the LSTM algorithm using Attention-CNN and LSTM.
Step 4	Steps 2 and 3 should be repeated until the loop convergence is reached.

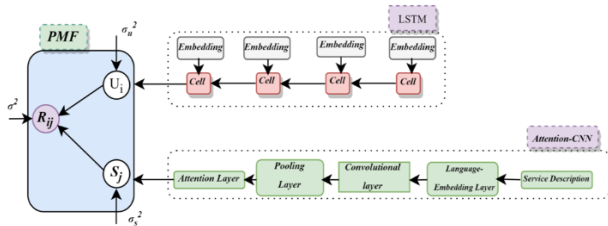


Fig. 1. DL-PMF model framework

The hidden feature vectors are detected by the Attention-CNN used from the service information. The DL-PMF model is part of a CNN-based network that is interested in this process. We were able to obtain accurate results based on the achievement of the natural language processing process for application programs to text emotion classification as well as analysis. Potential features can be easily extracted from global information using CNN [4, 6, 5, 7]. The focus is on each characteristic of the customer information description and the features of most of the API information description, in line with the visual approach and human information. These features and their meaning can also be obtained due to the in-depth study that is part of the attention process. The combination of CNN and the Focus Approach makes it possible to easily get better features as well as remove useless features from the Global Information filter. We will see in the proposed DL-PMF model how to retrieve hidden feature vectors using user information by LSTM with memory function. Current and past interests can be fulfilled based on history record information by remembering precise guarantees for customer perceptions and timing, and by remembering the implicit connection features between their respective APIs. The internal construction of Attention-CNN, as well as LSTM, is an overview as shown below.

A. The Structure of Attention-CNN Network

The Attention-CNN architecture is known as the network architecture seen in Fig. 2 and this architecture's main objective is to identify and acquire its tacit vectors based on the details contained in the document's service specification. Here, in the service text description, we convert the data into digital quantity and take it as input to the Attention-CNN model. Complex instances are correlated with word embedding and hence controllable measurements can be obtained. In this proposed model, we've introduced the embedding process.

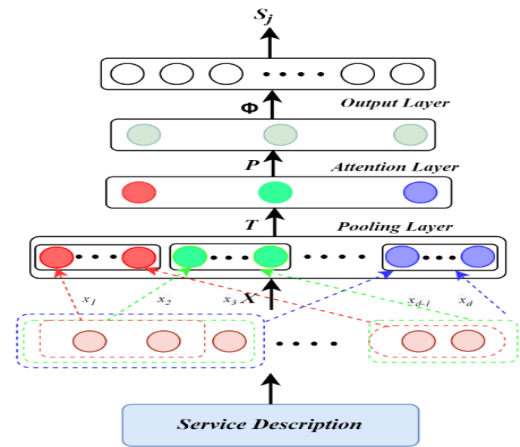


Fig. 2. Architecture of attention-CNN

1) *Language embedding layer*: The language embedding layer collects the raw material for the next level of the convolution layer and converts it into a dense numeric matrix and gives it as input. The word description of the service is used to input the words in the documents into  $d$  word vectors, converting the document into a matrix and input to the language embedding layer. Each word in the document is converted to a low-dimensional numeric vector using the pre-trained word2vec tool. We get the output of the layer as  $X \in R^{k \times d}$ , denoting the maximum length  $d$  of the vectors.

$$X = \begin{bmatrix} | & | & | & \dots & | & | & | & | \\ x_1 & x_2 & x_3 & \dots & x_{i-1} & x_i & x_{i+1} & \dots & x_d \\ | & | & | & \dots & | & | & | & | \end{bmatrix} \quad (2)$$

Where  $k$  is each word vector's dimension,  $x_i \in R^k$ .

2) *Convolution layer*: The key role of this layer has been described as gathering feature information related to the text. It is used to locate text attributes with different sharing weights using various convolution kernels. The length of the kernel window is then defined as  $l$ , and Eq. (3) is used to express the mutual weight matrix. The properties gained by the convolution kernel's layer  $j$  are expressed as follows.

$$t^j = [t_1^j, t_2^j, \dots, t_i^j, \dots, t_{d-l+1}^j] \quad (3)$$

The output eigenvalue obtained by a convolution kernel concerning region  $ix$  is denoted by  $t_i^j$ , and  $t_i^j = f(W^j \otimes X_{(i:(i+l-1))} + b^j)$ . The paper will specify the configuration operator as  $\otimes$ , the non-linear activation function as  $f(\cdot)$ , and the activation function as ReLU. From the  $i$  column of  $X$  to  $i + l - 1$  column by  $X_{(i:(i+l-1))}$ , and we identified the mapping matrix  $W^j$  and the bias by  $b^j$ .

This layer's outcome is:

$$T = [t^1, t^2, t^3, \dots, t^i, \dots, t^n] \quad (4)$$

Here,  $n$  is the kernel convolution number.

3) *Pooling layer*: This current layer captures the representative properties through the convolution layer on the



back, and constructs a fixed-length vector of equal length with the document variable following the pooling operation. The variable length obtained by the convolution layer represents a contextual feature vector with  $d - l + 1$ . This constant change in length makes it difficult to form subsequent layers. So we used max-pooling in this case. The maximum contextual characteristic of a piece of information is derived from each contextual feature vector that represents it and is represented by a shorter fixed-length vector. To substitute the output of that position in the network, this layer utilizes the general features of an adjacent output at a region. The layer's outcome is:

$$P = [p_1, p_2, \dots, p_i, \dots, p_n] \quad (5)$$

Where  $P_i = \max(t^i)$ ,  $\max(\cdot)$  is the maximum in  $t^i$ .

4) *Attention Layer*: At this stage, the layer has been introduced with more attention-grabbing approaches. The service weight is calculated from the information weight in each term based on the attention matrix approach. This not only highlights some of the important features but also leaves out unnecessary information using weight value in terms of size. So the corresponding weight for each word is calculated according to (6).

$$\alpha_i = g(W^{attn} * (p_i^{attn})^T + b^{attn}) \quad (6)$$

Where  $P_i^{attn} = [P_{i+\frac{-m+1}{2}}, P_{i+\frac{-m+3}{2}}, \dots, P_i, \dots, P_{i+\frac{m-1}{2}}]$ ,  $m$  is the magnitude of the sliding window.  $W^{attn} \in R^{l \times m}$ ,  $b^{attn} \in R$ .  $g(\cdot)$  is a function of non-linear activation.

The outcome of this current layer is

$$\Phi = [\omega_1, \omega_2, \dots, \omega_i, \dots, \omega_n] \quad (7)$$

Where  $\omega_i \in \alpha_i P_i$ .

5) *Output layer*: After obtaining high-level attributes from the previous layer in this sequence, they must be changed gradually for specific work as soon as the output layer is reached. That is, high-dimensional attributes are converted into a specific size feature vector and a latent attribute representation for customer service after reaching the layer before the output layer. The output to the output layer can be specified as follows.

$$s_j = \tanh(W_o \Phi + b_o) \quad (8)$$

From the above equation, we reference to  $W_o \in R^{D \times n}$  as the mapping matrix,  $b_o \in R^D$  as the bias, and  $s_j \in R^D$  as the service's latent characteristic vector  $j$ .

### B. LSTM Neural Network Structure

The correlation between customer and service features is very strong. This paper considers it a series of tasks to identify user characteristics for this correlation exploration. In this series of processes, we use the LSTM model to thoroughly search and dig into the correlations between expected features and to identify service users individually. Neural networks are specifically designed as "Recurrent Neural Networks (RNN)" and are used to process sequences of data. This allows each layer to be output to the next layer as well as its current state

for processing and RNNs are good at mining the interrelationships between cryptographic patterns. RNN showed better results in gradient disappearance and gradient explosion. Developed an evolutionary version of RNN called "Long Short-Term Memory (LSTM)" [37].

Another special type of RNN, LSTM, replaces the complex neural network structure by internally filling cells with low-cell neurons [38] Here, the LSTM of RNN is used to explore the properties of services that are interrelated through this framework and to independently evaluate the features of the services. Also, in evaluating subsequent characteristics, the LSTM internal structure indicates the secret state in the pre-history information and works better in the sequence modeling process. As input information, it takes complete sequence information, models it, and processes it into a single-dimensional vector. The inner component structure of the contributing LSTM-Cell can be described as shown in Fig. 3.

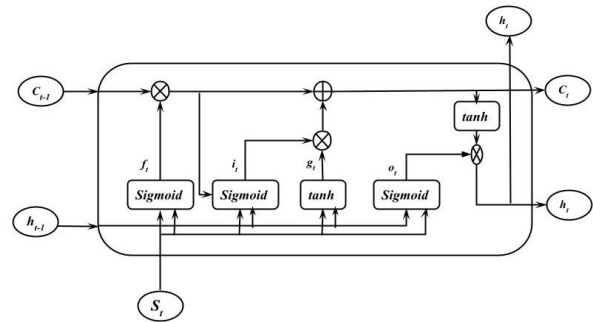


Fig. 3. Inner components of LSTM-Cell

Initially, LSTM, like RNN, considers service inputs to be interconnected and considered as a time series, and finally, its internal structure has become an excellent answer to the problem of bursting and disappearing gradient problems [16]. Below are the four significant aspects of the LSTM model. As shown in the figure: the input gate, forget gate, gate of the input module, and gate of the output. The gate and output gates of the module are used via the input gate to change the status of the cell, forget the maintenance gate, and control the deletion. The numerical method can be represented by the following equations for this procedure.

$$\left\{ \begin{array}{l} i_t = \text{sigmoid}(W_i * s_t + W_i * h_{t-1} + b_i) \\ f_t = \text{sigmoid}(W_f * s_t + W_f * h_{t-1} + b_f) \\ o_t = \text{sigmoid}(W_o * s_t + W_o * h_{t-1} + b_o) \\ g_t = \tanh(W_g * s_t + W_g * h_{t-1} + b_g) \\ C_t = f_t \cdot C_{t-1} + i_t \cdot g_t \\ h_t = \tanh(g_t) \end{array} \right. \quad (9)$$

Where the current cell status value, last time frame cell status value, and the update for the current cell status value are expressed by  $C_t$ ,  $C_{t-1}$ , and  $g_t$ . Input gate, forget gate, and output gate, respectively, are the notations  $i_t$ ,  $f_t$ , and  $o_t$ . With suitable input parameters, according to Eqs (9), the output value  $h_t$  is determined based on  $g_t$  and  $C_{t-1}$  values. Based on the difference between the output value and the actual value

after the back-propagation through time (BPTT) algorithm [39], all weights, including:  $W_i$ ,  $W_f$ ,  $W_o$ , and  $W_g$ , are modified.

This paper describes how the latent feature vector is derived from users' historical record information. The services used in the information in the usage history record over time need to be considered in the acquisition process and for this, the process of learning the latent feature vector can be considered as a time series approach. The answer to this approach is to use the LSTM architecture shown in Fig. 4 below to extract the latent feature vector from user information.

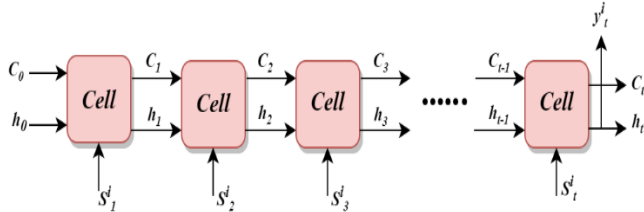


Fig. 4. LSTM External Architecture

The output and latent feature vectors of historical customer services can be obtained by input series, chronologically inserted into the LSTM model.

$$\left\{ \begin{array}{l} C_t = \sigma(W_1 \cdot [h_{t-1}, s_t^i] + b_1) * \tanh(W_2 \cdot [h_{t-1}, s_t^i] + b_2) + \\ \quad \sigma(W_3 \cdot [h_{t-1}, s_t^i] + b_3) * C_{t-1} \\ h_t = \sigma(W_4 \cdot [h_{t-1}, s_t^i] + b_4) * \tanh(C_t) \\ y_t^i = h_t \end{array} \right. \quad (10)$$

The mapping templates are denoted by  $W_1$ ,  $W_2$ ,  $W_3$ , and  $W_4$ , the bias is denoted by  $b_1$ ,  $b_2$ ,  $b_3$  and  $b_4$ , and the cell state by  $C_t$  at  $t$ , the latent attribute vector relative to the user  $i$  by  $y_t^i \in R^D$  and here  $\sigma(\cdot)$  is the activation function related to the sigmoid.

### C. Methodology for Proposed Model

In Fig. 1, the user  $i$ 's latent feature vector is denoted as  $U_i \in R^D$ , the latent attribute vector of the service  $j$  as  $S_j \in R^D$ , and the user  $i$  rating of the service  $j$  as  $R_{ij}$ . The variation of the Gaussian distribution of  $U_i$  and  $S_j$  is denoted by  $\sigma_U^2$  and  $\sigma_S^2$ , respectively.

The probability of a true conditional rating matrix [14] is determined as follows:

$$P(R|U, S, \sigma^2) = \prod_i^N \prod_j^M [N(R_{ij}|U_i^T S_j, \sigma^2)]^{I_{ij}} \quad (11)$$

The probability density of the Gaussian distribution with the  $\mu$ 's mean and  $\sigma^2$  variance is given by  $N(x|\mu, \sigma^2)$  suggested here. If the user  $i$  rates the service  $j$ , the index function  $I_{ij}$  value is set to 1, otherwise, its value is set to 0 will allocate. Assuming that the total number of users is  $N$  and the services are  $M$ , we identify the user's feature matrix by  $U \in R^{D \times N}$  and the latent feature matrix of services by  $S \in R^{D \times M}$ .

PMF is a sequence of consumers' and services' latent feature vectors that are independent of each other. Also, before the Gaussian distribution, the likelihood assumes the formula shown below for a zero-mean value.

$$(U|\sigma_U^2) = \prod_{i=1}^N N(U_i|0, \sigma_U^2 I) \quad (12)$$

$$P(U|\sigma_S^2) = \prod_{j=1}^M N(S_j|0, \sigma_S^2 I) \quad (13)$$

It is possible to substitute the Attention-CNN and LSTM outputs with equations (12) and (13).

$$P(U|W_U, X_U, \sigma_U^2) = \prod_{i=1}^N N(U_i|y^i, \sigma_U^2 I) \quad (14)$$

$$P(U|W_S, X_S, \sigma_S^2) = \prod_{j=1}^M N(S_j|s_j, \sigma_S^2 I) \quad (15)$$

In LSTM,  $W_U$  is used to denote the whole mapping matrix and its biases, and  $W_S$  is used to denote both mapping matrices with their biases. We can understand that the latent potential of the latent feature matrix for customers and services can be satisfied using the following formula, according to the Bayesian formula.

$$\begin{aligned} P(U, S, W_U, W_V|R, X, \sigma^2, \sigma_U^2, \sigma_S^2, \sigma_{W_U}^2, \sigma_{W_S}^2) \approx \\ P(R|U, S, \sigma^2) \times P(U|W_U, X_U, \sigma_U^2) \times P(W_U|\sigma_U^2) \times \\ P(S|W_S, X_S, \sigma_S^2) \times P(W_S|\sigma_S^2) \end{aligned} \quad (16)$$

For the prior probability distribution  $P(W_U|\sigma_{W_U}^2) = \prod_{\theta}^{\Omega} N(0, w_{\theta}|\sigma_{W_U}^2 I)$  relative to  $W_U$ , the total representation mapping matrix in the LSTM network can be observed as  $W_U$ . Priority probability distribution concerning  $W_S$  is  $P(W_S|\sigma_{W_S}^2) = \prod_{\phi}^{\mathfrak{R}} N(0, w_{\phi}|\sigma_{W_S}^2 I)$ , the total mapping matrix represented in Attention-CNN Can be observed as  $W_S$ . Also  $X_U$  in LSTM and  $X_S$  in Attention-CNN are referred to as the input set respectively.

In estimating the maximum posterior likelihood, the estimation matrix  $R$  found based on the DL-PMF model offered a nice solution.

$$\begin{aligned} \max \left( P(U, S, W_U, W_V|R, X, \sigma^2, \sigma_U^2, \sigma_S^2, \sigma_{W_U}^2, \sigma_{W_S}^2) \right) = \\ \max \left( P(R|U, S, \sigma^2) \times P(U|W_U, X_U, \sigma_U^2) \times P(W_U|\sigma_U^2) \times \right. \\ \left. P(S|W_S, X_S, \sigma_S^2) \times P(W_S|\sigma_S^2) \right) \end{aligned} \quad (17)$$

We consider the natural logarithm for the above equation for convenience. In

$$\begin{aligned} \left( P(U, S, W_U, W_V|R, X, \sigma^2, \sigma_U^2, \sigma_S^2, \sigma_{W_U}^2, \sigma_{W_S}^2) \right) = \\ -\frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{j=1}^M I_{ij} (R_{ij} - U_i^T S_j)^2 - \frac{1}{2\sigma_U^2} \sum_{j=1}^M \|S_j - \\ s_j\|^2 - \frac{1}{2\sigma_U^2} \sum_{i=1}^N \|U_i - y^i\|^2 - \frac{1}{2\sigma_{W_U}^2} \sum_{\theta}^{\Omega} \|w_{\theta}\|^2 - \\ \frac{1}{2\sigma_{W_S}^2} \sum_{\phi}^{\mathfrak{R}} \|w_{\phi}\|^2 + C \end{aligned} \quad (18)$$

The LSTM, Attention CNN mapping matrix numbers, are denoted as  $\Omega$  and  $\mathfrak{R}$  respectively, and the matrix entries in the equation (16) can be omitted since the LSTM, Attention CNN mapping matrices, are constructed randomly according to the distribution basis of Gaussian.

To make the maximum value of equation (18) equal to the minimum, the following equation is used:

$$\begin{aligned} f = \sum_{i=1}^N \sum_{j=1}^M I_{ij} (R_{ij} - U_i^T S_j)^2 + \chi \sum_{j=1}^M \|S_j - s_j\|^2 + \\ \beta \frac{1}{2\sigma_U^2} \sum_{i=1}^N \|U_i - y^i\|^2 \end{aligned} \quad (19)$$

Where  $\chi = \sigma^2/\sigma_s^2$ ,  $\beta = \sigma^2/\sigma_u^2$ .

The algorithm of coordinate descent is used in this paper to find and solve the minimum value of  $f$ . The  $U_i$  and  $S_j$  updates are expressed as follows:

$$U_i \leftarrow \left( 2(2\beta I_k + \sum_{j=1}^M I_{ij} S_j S_j^T) \right)^{-1} * (\beta * y^i + \sum_{j=1}^M R_{ij} I_{ij} S_j) \quad (20)$$

$$S_j \leftarrow \left( 2(2\chi I_k + \sum_{i=1}^M I_{ij} U_i U_i^T) \right)^{-1} * (\beta * v_j + \sum_{i=1}^M R_{ij} I_{ij} U_i) \quad (21)$$

## V. EXPERIMENTAL EVALUATION

Initially, each value of the matrix in the process results indicated here is denoted by 0 or 1 and equated with the performance of the four other methods. In this case, any API followed by the user is marked as '0' and otherwise as '1'.

### A. Datasets

We used as a package the existing API data from the website called *programmableweb* and the API dataset was scrambled with comprehensive information about them. We scrambled a count of 17412 API data for this data set. It includes a list of followers for each API, as well as developer information, a quick summary, a range of names, and a post date. And here we have extracted the new follow feature from the web along with customer information. Consumer information description is considered equivalent to the performance of the goods. As followers of 17412 APIs, we have compiled the data of 140,000 users for this purpose and we have included the details in Table II.

TABLE II. STATISTICS FROM CRAWLED DATASET

<i>Scrambled Dataset</i>	<i>Total (#)</i>
User profiles	17412
API's	22032
Invocation files	248530

All the data in the scrambled dataset is split into training, validation, and test sets to evaluate the results of the experiments we perform. Random selection of training sets to 90%, 80%, 70%, 60%, and 10% is possible due to the arrangement of the data settings. In this experiment, 90 data sets for training, 5 for validation, and 5 percent for test density were analyzed first, and for the other four data settings for 10% validation of the data in the data set, the remaining data for the test set was composed. I.e. training, certification, and test sets with 10%, 10%, and 80% concentrations were taken. Thus experiments were performed ten times for each group. We analyzed all the results for each group experiment with MAE (Mean Absolute Error) and RMSE (Root-mean-square error) values.

### B. Evaluation Metrics

There has been an introduction to a recommended model to estimate a type. The goal is to determine the prediction accuracy of the API suggestion estimates based on the trials

completed. We used the MAE and RMSE as evaluation criteria.

$$MAE = \frac{1}{N} \sum_{i,j} |R_{ij} - \hat{R}_{ij}| \quad (22)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i,j} (R_{ij} - \hat{R}_{ij})^2} \quad (23)$$

N stands for the numeral of tests,  $R_{ij}$  stands for the user's rating of service  $j$ , and  $\hat{R}_{ij}$  stands for the user's estimated rating of service  $j$  using the stated method.

### C. Parameter Setting

Since the model, we proposed above depends on the latent factor method (LFM), this model is particularly affected by two parameters. The first is the total quantity of hidden features: though its value is too high it cannot be understood by the model. Yet if this value is too low, we will get negative effects from the model. So, we took this number to be 10. The second introduced a parameter called regularization  $\lambda$  to adjust the weight of the term. As part of this,  $\lambda$  is set to 0.1, as well as the regularization value  $\lambda_u, \lambda_a$  is set to 0.01 related to user terms and related API terms.

### D. Comparative Result Evaluation of Performance

To evaluate the effectiveness of the proposed strategy, we used the following methods and the experimental findings are listed in below Table III.

TABLE III. EXPERIMENT RESULTS

Method	The density of the Training set (T)									
	T=90%		T=80%		T=70%		T=60%		T = 10%	
	MAE	RMS E	MAE	RMS E	MAE	RMS E	MAE	RMS E	MAE	RMS E
PNF	0.39		0.41		0.42		0.44		0.63	
	1	0.555	1	0.571	6	0.584	1	0.596	5	0.739
NMF	0.26		0.28		0.31		0.39		0.40	
	5	0.531	3	0.564	5	0.623	5	0.747	2	0.818
ConvMF	0.46		0.46		0.47		0.48		0.48	
	8	0.684	9	0.685	1	0.69	1	0.696	2	0.719
Autoencoder	0.29		0.33		0.34		0.37		0.41	
	9	0.378	2	0.378	4	0.387	7	0.377	7	0.45
DL-PMF	0.15		0.16		0.17		0.19		0.42	
	1	0.204	3	0.223	6	0.246	3	0.276	4	0.646

1) *PMF* [40]: For collaborative filtering, Probabilistic Matrix Factorization incorporates ratings.

2) *NMF (Non-Negative MF)* [41]: A Non-Negative matrix factorization is modeled as the product of the two non-negative matrices.

3) *ConvMF* [7]: CNN is used to extract item information in Convolutional Matrix Factorization.

4) *Autoencoder* [42]: A multilayer neural network that is commonly used in feature learning to reduce dimensionality.

5) *DL-PMF*. Our method combines the PMF with a CNN and an LSTM network to predict recommendation results.

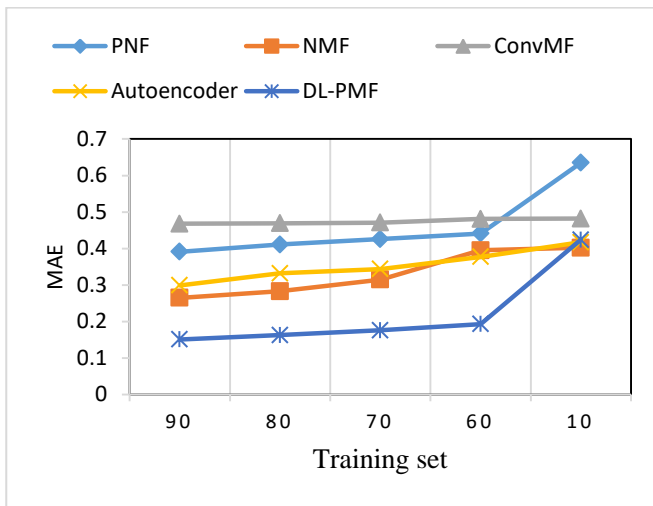


Fig. 5. Impact on MAE

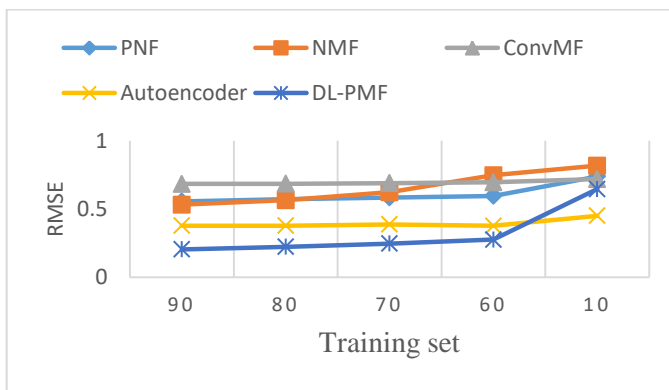


Fig. 6. Impact on RMSE

We have shown the line chart as shown in Fig. 5 and Fig. 6, clearly describing the model proposed by us by comparing the working methods of different methods with training sets of different densities and describing its typical performance. Fig. 5 and Fig. 6 show the MAE and RMSE values for training sets of different densities. The consistent performance of the test set with relatively different densities is maintained in the NMF approach. So, depending on this it is understood that the logistic regression here is relatively constant. The difference is huge. MAE and RMSE values are higher in the autoencoder model than in ConvMF when the concentration of the test set is 90%. In general, the value of this difference is much lower in the model proposed by us and works better than in any other algorithm. In IoT, our approach will deliver stronger API recommendations not only in terms of data efficiency but also in terms of perceiving features related to large data obtained in the IoT system.

## VI. CONCLUSION

The DL-PMF algorithm purposed for mentioned paper enhances the efficiency of the recommendation model by perceiving the implicit nature of the users and API elements. For this purpose, the model of the API treats only relevant information by filtering non-interfering information in the description. Our launch results have demonstrated that obtaining the latent characteristics of users and API elements

effectively through our proposed DL-PMF model can significantly raise the recommendation's accuracy. In this instance, the recommendation accuracy will not only be considered as contextual information but in future work, we will consider the collection of features in the description information by RCNN.

## REFERENCES

- [1] T. Ramathulasi and M. R. Babu, "Comprehensive Survey of IoT Communication Technologies," In Emerging Research in Data Engineering Systems and Computer Communications, pp. 303-311, 2020.
- [2] S. Zhang, L. Yao, A. Sun and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," ACM Computing Surveys (CSUR), vol. 52, no. 1, pp. 1-38, 2018.
- [3] Y. Koren, "Factorization meets the neighborhood: a multifaceted collaborative filtering model.," in In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 2008, August.
- [4] L. Zheng, V. Noroozi and P. S. Yu, "Joint deep modeling of users and items using reviews for recommendation," in In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, 2017.
- [5] H. Wang, N. Wang and D. Y. Yeung, "Collaborative deep learning for recommender systems.," in In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, 2015.
- [6] D. Kim, C. Park, J. Oh, S. Lee and H. Yu, "Convolutional matrix factorization for document context-aware recommendation.," In Proceedings of the 10th ACM conference on recommender systems, pp. 233-240, 2016.
- [7] D. Kim, C. Park, J. Oh and H. Yu, "Deep hybrid recommender systems via exploiting document context and statistics of items," Information Sciences, 417., pp. 72-87, 2017.
- [8] X. Shen, B. Yi, Z. Zhang, J. Shu and H. Liu, "Automatic recommendation technology for learning resources with convolutional neural network.," In 2016 International Symposium on Educational Technology (ISET), IEEE., pp. 30-34, 2016.
- [9] S. Okura, Y. Tagami, S. Ono and A. Tajima, "Embedding-based news recommendation for millions of users," in the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017.
- [10] B. H. Devooght R, "Collaborative filtering with recurrent neural networks," arXiv preprint arXiv:1608.07400., 2016.
- [11] S. Seo, J. Huang, H. Yang and Y. Liu, "Representation Learning of Users and Items for Review Rating," International Workshop on Machine Learning Methods for Recommender Systems (MLRec)(SDM'17)., 2017.
- [12] J. Chen, H. Zhang and X. He, "Attentive collaborative filtering: Multimedia recommendation with," Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval. ACM, pp. 335-344, 2017.
- [13] T. Ramathulasi and M. Rajasekhar Babu, "Enhanced PMF Model to Predict User Interest for Web API Recommendation.," in P. Krishna (Ed.), Handbook of Research on Advances in Data Analytics and Complex Communication Networks, 2022.
- [14] B. Twardowski, "Modelling contextual information in session-aware recommender systems with neural networks.," in In Proceedings of the 10th ACM Conference on Recommender Systems. ACM, 2016.
- [15] H. Soh, S. Sanner, M. White and G. Jamieson, "Deep sequential recommendation for personalized adaptive user interfaces," in Proceedings of the 22nd International Conference on Intelligent User Interfaces, 2017.
- [16] L. Yao, X. Wang, Q. Z. W. Shenh, Ruan and W. Zhang, "Service recommendation for mashup composition with implicit correlation regularization," in Int. Conf. Web Serv. (ICWS), 2015.

- [17] Z. Zheng, H. Ma, L. Lyu and I. King, "WSRec: A collaborative filtering based web service recommender system," in IEEE Int. Conf. Web Serv., 2009.
- [18] Z. Zheng, H. Ma, M. Lyu and I. King, "Collaborative web service QOS prediction via neighborhood integrated matrix factorization," IEEE Trans. Serv. Comput., vol. 6, p. 289–299, 2013.
- [19] B. Cao, J. Liu, M. Tang, Z. Zheng and G. Wang, "Mashup service recommendation based on user interest and social network," in IEEE Int. Conf. Web Serv. (ICWS), 2013.
- [20] H. Xue and D. Zhang, "A recommendation model based on content and social network," in In 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2019, May.
- [21] X. Wang, H. Wu and C. H. Hsu., "Mashup-oriented api recommendation via random walk on knowledge graph," IEEE Access, vol. 7, pp. 7651–7662, 2018.
- [22] F. Xie, L. Chen, D. Lin, C. Chen, Z. Zheng and X. Lin, "Group preference based API recommendation via heterogeneous information network," in In Proceedings of the 40th International Conference on Software Engineering: Companion Proceedings, 2018.
- [23] S. P. Ma, H. J. Lin, C. A. Yu and C. Y. Lee, "Web API recommendation based on service cooperative network," in In 2017 International Conference on Applied System Innovation (ICASI), 2017, May.
- [24] H. Sun, Z. Zheng, J. Chen, W. Pan, C. Liu and W. Ma, "Personalized open API recommendation in clouds via item-based collaborative filtering," in In 2011 Fourth IEEE International Conference on Utility and Cloud Computing, 2011, December.
- [25] K. Fletcher, "Regularizing matrix factorization with implicit user preference embeddings for web API recommendation," in In 2019 IEEE International Conference on Services Computing (SCC), 2019, July.
- [26] Y. Chen, M. Zhou, Z. Zheng and D. Chen, "Time-aware smart object recommendation in social internet of things," IEEE Internet of Things Journal, vol. 7, no. 3, pp. 2014–2027, 2019.
- [27] S. Duan, D. Zhang, Y. Wang, L. Li and V. Zhang, "JointRec: A deep-learning-based joint cloud video recommendation framework for mobile IoT.," IEEE Internet of Things Journal, vol. 7, no. 3, pp. 1655–1666, 2019.
- [28] Z. Huang, X. Xu, J. Ni, H. Zhu and C. Wang, "Multimodal representation learning for recommendation in Internet of Things," IEEE Internet of Things Journal, vol. 6, no. 6, pp. 10675–10685, 2019.
- [29] H. X. Hu, B. Tang, Y. Zhang and W. Wang, "Vehicular ad hoc network representation learning for recommendations in internet of things," IEEE Transactions on Industrial Informatics, vol. 16, no. 4, pp. 2583–2591, 2019.
- [30] T. Ramathulasi and M. Rajasekharababu, "Augmented latent Dirichlet allocation model via word embedded clusters for mashup service clustering," Concurrency and Computation: Practice and Experience, vol. 34, no. 15, p. e6896, July 2022.
- [31] T. Ramathulasi and M. R. Babu, "Enhancing Clustering Performance Using Topic Modeling-Based Dimensionality Reduction," International Journal of Open Source Software and Processes (IJOSSP), vol. 13, no. 1, pp. 1–16, 2022.
- [32] Y. Shen, X. He, J. Gao, L. Deng and G. Mesnil, "A latent semantic model with convolutional-pooling structure for information retrieval," In Proceedings of the 23rd ACM international conference on conference on information and knowledge management, pp. 101–110, 2014.
- [33] Y. Kim, "Convolutional neural networks for sentence classification," In Proceedings of the 2014 Empirical Methods in Natural Language Processing (EMNLP), p. 1746–1751, 2014.
- [34] R. Collobert, J. Weston, M. Bottou and K. Karlen, "Natural language processing (almost) from scratch," Natural language processing (almost) from scratch. Journal of Machine Learning Research (JMLR), p. 2493–2537, 2011.
- [35] A. Van Den Oord, S. Dieleman and B. Schrauwen, "Deep content-based music recommendation".
- [36] Y. Hu, Y. Koren and C. Volinsky, "Collaborative Filtering for Implicit Feedback Datasets".
- [37] S. Hochreiter and J. Schmidhuber, "Long short-term memory. Neural computation, 9(8)," pp. 1735–1780., 1997.
- [38] R. Jozefowicz, W. Zaremba and I. Sutskever, "An empirical exploration of recurrent network architectures," In International conference on machine learning, pp. 2342–2350, 2015, June.
- [39] P. J. Werbos, "Backpropagation through time: what it does and how to do it.," Proceedings of the IEEE, 78(10), pp. 1550–1560, 1990.
- [40] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," In Advances in neural information processing systems, pp. 1257–1264, 2008.
- [41] M. N. Schmid, O. Winther and L. K. Hansen, "Bayesian non-negative matrix factorization," In International Conference on Independent Component Analysis and Signal Separation, pp. 540–547, 2009, March.
- [42] J. Liu, Y. Qiu, Z. Ma and Z. Wu, "Autoencoder based API Recommendation System for Android Programming," In 2019 14th International Conference on Computer Science & Education (ICCSE), pp. 273–277, 2019, August.

# Unsupervised Learning-based New Seed-Expanding Approach using Influential Nodes for Community Detection in Social Networks

Khaoula AIT RAI<sup>1</sup>, Mustapha MACHKOUR<sup>1</sup>, Jilali ANTARI<sup>2</sup>

Computer System and Vision Laboratory-Faculty of Sciences, Ibn Zohr University, Agadir BP8106, Morocco<sup>1</sup>  
Laboratory of Computer Systems Engineering-Mathematics and Applications-Polydisciplinary Faculty of Taroudant, Ibn Zohr University, Morocco<sup>2</sup>

**Abstract**—Several recent studies focus on community structure due to its importance in analyzing and understanding complex networks. Communities are groups of nodes highly connected with themselves and not much connected to the rest of the network. Community detection helps us to understand the properties of the dynamic process within a network. In this paper, we propose a novel seed-centric approach based on TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) and k-means algorithm to find communities in a social network. TOPSIS is used to find the seeds within this network using the benefits of multiple measure centralities. The use of a single centrality to determine seeds within a network, like in classical algorithms of community detection, doesn't succeed in the majority of cases to reach the best selection of seeds. Therefore, we consider all centrality metrics as a multi-attribute of TOPSIS and we rank nodes based on the TOPSIS' relative closeness. The Top-K nodes extracted from TOPSIS will be considered as seeds in the proposed approach. Afterwards, we apply the k-means algorithm using these seeds as starting centroids to detect and construct communities within a social network. The proposed approach is tested on Facebook ego network and validated on the famous dataset having the ground-truth community structure Zachary karate club. Experimental results on Facebook ego network show that the dynamic k-means provides reasonable communities in terms of distribution of nodes. These results are confirmed using Zachary karate club. Two detected communities are detected with higher normalized mutual information NMI and Adjusted Rand Index ARI compared to other seed centric algorithms such as Yasca, LICOD, etc. The proposed method is effective, feasible, and provides better results than other available state-of-the-art community detection algorithms.

**Keywords**—Complex network; community detection; TOPSIS; seed-centric approach; ground-truth; k-means

## I. INTRODUCTION

Many complex real-world systems can be represented and studied as networks. Complex networks cover diverse networks as the Internet, metabolic networks, social networks, and many others. Studies conducted on the physical significance and mathematical properties of complex networks have found that these networks share macroscopic properties. Among these properties, we cite prototype properties such as the small-world effect [1] and the free-scale [2], dynamic properties such as diffusion [3][4] and structural properties

such as community structure. The community structure property appears to be common to many complex networks and helps to understand the relationship between a single node in microscopy and groups in macroscopy. Communities are defined as parts of the network with numerous internal connections but few exterior connections. They are closely related to the functional components of real-world networks, such as metabolic networks cycles and pathways and protein complexes in protein-protein interaction networks. They can have very different topological properties than the whole network and then affect the dynamic of the network. Therefore, community structure discovery has been the focus of several recent efforts. Numerous approaches have been suggested to detect community structures in networks, some are based on similarity measures, and others rely on network dynamics such as random walk dynamics [5] and label propagation [6]. Other approaches rely on statistical models end on the optimization of quality functions. For instance, the well-known Newman-Girvan modularity [7] can be used as a method of community discovery and as an objective technique to measure the quality of community partitions.

In this paper, we propose a new approach to discover communities within a social network. The proposed approach uses sequentially two techniques. At the beginning, we apply TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) methodology [8] which aggregates centrality measures (degree centrality [9], betweenness centrality [10], closeness centrality [9], and eigenvector centrality [11]) as multi-attribute to rank nodes in a complex network, and then we run the k-means algorithm which is an unsupervised learning algorithm widely used for data clustering. To take advantage of this algorithm of clustering, we use the output of TOPSIS as initial centroids for k-means to get communities that may be in a complex network. To do so, we need to initialize the value of k by the desired output number of clusters. We did this in two ways; classical k-means and dynamic k-means using Elbow [12] and silhouette [13] methods. TOPSIS is a famous technique of multi criteria decision analysis. It's used in different fields such as Supplier selection [14], selecting the best wood type [15], selecting techniques for future avian influenza surveillance in Canada [16], personnel selection [17], etc.

The main contributions of this work are as follows:

- Taking advantage of different indexes of centrality gathered in TOPSIS to identify influential nodes.
- Based on top-K important nodes extracted from TOPSIS we propose a seed centric algorithm based on k-means to discover communities in complex networks. Experiments are realized using adjacency matrix as centrality measures of top-K influential nodes.
- To determine the optimal K we use Elbow and Silhouette methods and then applying the proposed approach with the optimal value of k.

This paper is then organized as following. Section II presents a general overview on seed centric algorithms. Section III details the general context and related concepts. Section IV explains the proposed approach. Section V presents the analysis and results of the experiments. The paper is concluded in Section VI.

## II. SEED CENTRIC ALGORITHMS: GENERAL OUTLINES AND RELATED WORKS

The algorithm of seed centric approaches is based on three important phases that are: the calculation of seeds, the calculation of seed local community, and community calculation from the previous step. In community detection field, there are several seed-based algorithms proposed by researchers. These algorithms are classified according to several factors. For the first phase for instance, we can find approaches that use single seed or group of seeds, linked or not [18]. Apart from the nature of the seeds (single or group)[19] [20], the number of seeds is also a factor that differs from an algorithm to another. There are some algorithms that have the number of seeds as an input, like the application of the classical k-means algorithm. Other algorithms use heuristic approaches to compute adequate seeds, for example, the approaches proposed by D. Shah et al. [18], [21] and Kanawati [18] based on leaders. D.Shah et al. [22] proposed two algorithms; Leader Follower Algorithm (LFA) and Fast Leader Follower Algorithm (FLFA). These algorithms find leaders as seeds of the community, and then they search its other members. During this process of leaders' search, they consider leaders as nodes with lower degree than their followers. Although the FLFA is fast, LFA can detect more communities in some networks. LICOD algorithm proposed by Kanawati [18] is also based on identifying leaders in the network then affecting the remaining nodes to these leaders to build communities. LICOD calculates the number of communities to detect automatically.

The selection of seeds can be random or informed. Random selection means choosing randomly adequate seeds with repetitive process while informed selection consists of choosing a set of nodes or subgraphs. LICOD algorithm proceeds by informed selection. It selects nodes with higher centrality.

For calculating seed local community, two approaches are applied by researchers in this area; expanding approach and agglomerative approach.

Expanding approach [23] relies on ego-centred algorithms for community detection. The limit of this approach is that it does not take into account all of the network's nodes. To overcome this problem, Whang et al. [24] proposed to attach outliers to the nearest community. The second approach for calculating seed local community is the agglomerative approach, where nodes are agglomerated into communities with nearest identified seed [18] [21].

The last common phase of all community detection algorithms is community calculation which leads to the final communities based on local seeds communities calculated in the precedent step. Fig. 1 illustrates the link between these approaches and algorithms.

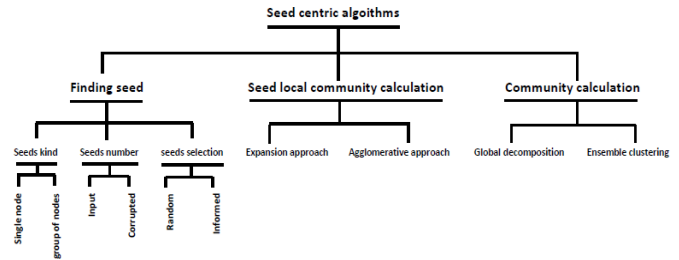


Fig. 1. Summary of seed centric algorithms' characteristics

In the last decade, several seed centric algorithms for community discovery have been proposed in the scientific literature. Each one relies on different seeds finding and different approaches for community calculation. Kanawati et al. [18] proposed Licod algorithm based on two nodes in the community: leaders and followers. Leaders are nodes with high centrality, whatever the centrality is. Followers are founded by computing community membership. Weskida [25] also search in this field how to select seeds using some evolutionary algorithms. The problem in this kind of algorithms is when the evaluation functions are noisy, they will not work. Yufeng Wang et al. [26] proposed an algorithm that selects seeds using page rank-like algorithm. Kanawati proposed also the Yasca algorithm that computes partitions in graphs using local community identification [27]. This algorithm applies an ensemble clustering to the local communities detected for each seed. Akrowska and Bader proposed also a seed set expansion algorithm of dynamic greedy [28]. In each iteration, the algorithm updates the local communities from an initial partition that is obtained via a set seed expansion method. Table I highlights several seed-centric algorithms' properties.

TABLE I. PROPERTIES OF SOME SEED CENTRIC ALGORITHMS [29]

Algorithm	Seed Nature	Seed number	Seed selection	Local community
Licod [18]	Set	Computed	Informed	Agglomerative
Yasca [27]	Single	Computed	Informed	Expansion
[30]	Subgraph	Computed	Informed	Expansion
[24]	Single	Computed	Informed	Expansion
[31]	Subgraph	Computed	Informed	Expansion
[32]	Set	Computed	Automatic	Expansion

### III. GENERAL CONTEXT

In this section, we present fundamental concepts that constitute the general context of the proposed approach. These concepts complete each other to emphasize the proposed idea.

#### A. TOPSIS Methodology

The Technique for Order Preference by Similarity to the Ideal Solution (TOPSIS) [8] emerged in the 1980s as a decision-making method based on several criteria. It chooses the alternative of the shortest Euclidean distance from the ideal solution and the longest distance from the negative ideal solution. TOPSIS demonstrates its power to solve this MCDM (Multi-Criteria Decision Making) problem in different fields such as supply chain management, engineering, health, design, etc. TOPSIS has gained considerable interest from the scientific community due to its success in various areas. TOPSIS was applied for supplier selection [33]. Different factors are considered by Chen et al [33] like quality and technological power. Yong used TOPSIS also to choose the plant location [34]. In the field of human resource management, Kelemenis and Askounis [17] used the fuzzy TOPSIS to select the best management member in an IT department. Wang and Elhag [35] used Fuzzy TOPSIS and nonlinear programming for selecting a system analysis engineer. Kaya et Kahraman [36] applied fuzzy TOPSIS combined with fuzzy AHP to Select the best energy technology alternative.

The process of TOPSIS consists of the following main steps [8]:

**Step 1:** construction of the normalized decision matrix

**Step 2:** construction of the weighted normalized decision matrix

**Step 3:** determination of the positive and the negative ideal solutions

**Step 4:** calculation of the separation of each alternative

**Step 5:** the calculation of the relative closeness to the ideal solution.

How each step is modeled and used is explained in the Section III.

#### B. Centrality Measures

The computation of centrality measures has been an important issue in the field of social network analysis for several decades [37]. Centrality is a notion that makes it possible to account the popularity or the visibility of an element within a group. Freeman's article "Centrality in social networks: Conceptual clarification [9]" represents arguably one of the most important contributions in the field of social network analysis. In his article, Freeman proposes three formal definitions of the concept of centrality that we present below. We also present a fourth centrality measure introduced by Bonacich.

1) *Degree centrality*: It represents the basic and the most concise way of the notion of centrality. It is based on the concept that a person's influence within a group depends on the total number of individuals he knows or has direct contact

with [9]. According to this measurement, indicating the value of a node in a graph depends on the number of its neighboring vertices, i.e. the number of its incident links. In graph theory, Degree Centrality originally comes from this number, which is known as the node's degree.

Because it simply considers a node's immediate neighborhood and ignores the overall structure of the network, degree centrality is also known as local centrality measure [38]. While degree centrality is pertinent in some contexts, it is ineffective in others, such as the analysis of web page graphs [39].

2) *Closeness centrality*: It is a measure of global centrality based on the assumption that a node occupies a strategic (or favorable) position in a graph if it is globally nearby to the other nodes of this graph [9]. For example, in a social network, this metric refers to the fact that an actor is influential if he can quickly get in touch with a lot of other actors while exerting the least amount of effort (the effort here is relative to the size paths).

3) *Betweenness centrality*: It is another metric of global centrality that Freeman has presented [9]. The idea behind this measurement is that a node in a graph is important if a maximum number of other nodes cross it. More precisely, a node with a strong betweenness centrality is a node through which passes a large number of geodesic paths (i.e. shortest paths) in the graph. In a social network, an actor with a strong betweenness centrality is a node on which depends a large number of interactions between non-adjacent nodes [40]. In a communication network, betweenness centrality of a node can be considered as the probability that information transmitted between two nodes passes through this intermediate node [40].

4) *Eigenvector centrality*: It's a measure suggested by Bonacich[11] based the idea that the centrality of a node is determined by the centrality of the nodes to which it is connected. In a social network, this refers to the concept that an actor is influential when he is linked to other influential actors. In fact, it is an extension of degree centrality in which the same weight is not given to the neighboring nodes. Practically talking, Bonacich proposes to consider the centrality of a node as being dependent on the linear combination of the centralities of its neighboring nodes[11].

#### C. K-means Algorithm

K-means is an unsupervised algorithm widely used in data clustering. It proceeds by analyzing a set of data characterized by a set of descriptors, in order to group "similar" data into groups (or clusters).

To split a dataset into k distinct clusters, k-means algorithm needs a way to compare the degree of similarity between all the observations. Usually, the distance between two elements to compute their similarity is used.

Thus, two similar data will have a small dissimilarity distance, while two different data will have a big separation distance. The famous used metric to measure such similarity is the Euclidean distance [41], and it is the one used in this paper.



1) *Euclidian distance*: It's a geometric distance that considers a matrix  $X$  with  $n$  quantitative variables in the vector space  $E^n$ . The Euclidean distance  $d$  between two observations  $x_1$  and  $x_2$  is calculated as follows:

$$d(x_1, x_2) = \sqrt{\sum_{j=1}^n (x_{1n} - x_{2n})^2} \quad (1)$$

Algorithm 1 shows the principal of k-means algorithm. The beginning is the choice of  $k$  elements chosen arbitrarily from the dataset as centroids. Then the distance between all the elements and each one of these centroids is computed. Each element belongs then to the cluster whose centroid is closest to it. As a second step, new centroids are computed as the mean of all the elements of each cluster. These steps are repeated until there is no new centroid.

The purpose of clustering algorithms in general and K-Means specifically is to form clusters of similar elements. As long as they are stored in a data matrix, these items can be any kind of thing.

2) *K-means principal*: The sum of the distances between each item and the centroid is minimized by the iterative algorithm k-means. The outcome depends on the original selection of centroids.

Adopting a cloud of a given set of points, K-Means updates the members of each cluster until the sum can no longer reduce. Depending on picking the right value  $K$  for the number of clusters, the outcome is a collection of compact and separate clusters.

```
Algorithm 1: K-means algorithm
INPUT
    K: number of clusters to construct
    The training set
BEGIN
    Randomly choose K points from the dataset that will be
    the centroids of the starting clusters
REPEAT
    Assign each point (element of the dataset) to the cluster
    to which it is closest
    Update the centroid of each cluster by the mean of its
    points.
UNTIL convergence OR stabilization of total population
    inertia
END
```

The k-means algorithm may converge under one of the following cases:

- When the number of iterations is fixed in advance, K-means will run its iterations and then end, regardless of how the compound clusters are shaped.
- Stabilization of cluster centers (centroids no longer move during iterations).

The main challenge of k-means algorithm is the value of  $k$ . Indeed, it is not always evident to choose  $k$  as the number of clusters; particularly, if the dataset is sizable, and we don't

have a priori assumptions on the data. A big value of  $k$  can generate too fragmented partitioning data. This will prevent discovering interesting patterns in the data. That is on one hand; on the other hand, too small value of  $k$  will potentially generate general clusters containing a lot of data. In this case, there will be no "fine" patterns to discover. We need then to know the optimal value of  $k$ . The most used methods for this purpose are the Elbow and the Silhouette methods and the technique is called dynamic k-means.

#### D. Dynamic k-means

The challenge of any clustering algorithm is to determine the optimal number of clusters in which the data can be grouped. For the k-means, Elbow and Silhouette methods are the most popular methods for determining this optimal  $k$  value.

1) *Elbow method*: The Elbow Method [12] is one of the most popular methods that helps to find the optimal number of groups to which the k-means algorithm splits the dataset. The idea starts by varying the number of clusters  $k$  from 1 to  $N$ , assuming that the data has already been divided into  $k$  clusters by a clustering method. For each value of  $k$ , the WCSS (Within-Cluster Sum of Square) which is the sum of the squared distance between each point and the centroid of a cluster is calculated. The plot generated then in the visualization looks like an elbow. That's why it's nomination. Fig. 2 shows an example of the relation between WCSS values and clusters' number. While the WCSS value is decreasing, the value of  $k$  is increasing. The X-axis value where the curve appears as if it starts to bend represents the optimal value of  $k$ . This deformation value represents the elbow of the curve.

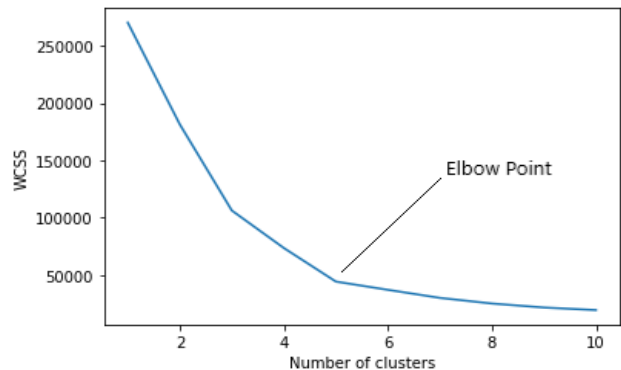


Fig. 2. Example of WCSS plot

2) *Silhouette method*: The silhouette algorithm [13] is usually used to find the optimal number of clusters for an unsupervised learning technique. In the Silhouette algorithm, we suppose that data has already been split into  $k$  groups by a clustering technique (typically k-means algorithm). It proceeds by computing silhouette coefficient for each data point to measure how well this point is assigned to the most appropriate cluster. Its value varies between -1 and 1. A value close to 1 means that the sample is assigned to the best cluster and vice versa. To compute the silhouette coefficient, the following values is needed:

$C(i)$ : the cluster assigned to the  $i^{\text{th}}$  data point.

$|C(i)|$ : the number of data points in  $C(i)$

$a(i)$  : gives a measure of the quality of the assignment of the  $i$ th data point to its cluster.

$$a(i) = \frac{1}{|C(i)|-1} \sum_{C(i), i \neq j} d(i, j) \quad (2)$$

$b(i)$ : It is defined as the average dissimilarity with the nearest cluster.

$$b(i) = \min_{i \neq j} \left( \frac{1}{|C(j)|} \sum_{j \in C(j)} d(i, j) \right) \quad (3)$$

The silhouette coefficient  $s(i)$  is then given by:

$$s(i) = \frac{b(i)-a(i)}{\max(a(i), b(i))} \quad (4)$$

Equation (4) gives the silhouette coefficient for each value of  $k$ , and the  $k$  having the maximum value of  $S(i)$  is taken as the optimal number of clusters for the unsupervised learning algorithm. Fig. 3 illustrates the plot corresponding to the silhouette coefficient's value for each value of  $k$ . From that plot we notice that the optimal value of  $k$  is 4.



Fig. 3. Silhouette coefficient's plot (maximum value of  $S(i)$  is 4)

#### IV. PROPOSED APPROACH

The main goal of the proposed approach is to detect communities in a social network. It's a task of clustering the dataset that requires the application of an appropriate algorithm. The famous algorithm used for clustering is k-means. Since this algorithm requires starting with a value of  $k$ , the main idea here is to use the top-  $K$  influential nodes within the used dataset. These nodes are detected using TOPSIS methodology and then are used as centroids to run k-means algorithm. These steps are synthesized in Algorithm 2 and Fig. 4.

##### Algorithm 2: Steps of the proposed algorithm

**Input:**  $G(V, E)$  : A social network,  
 $K$  : the number of influential nodes (number of communities),  
 $k = \{k_1, k_2, \dots, k_n\}$  : a set of centrality measures,  
 $w$  : weight given centralities.

**Output:** detected communities

**Begin**

1. Select the top- $K$  influential nodes  $P$  using TOPSIS:  
 $P \leftarrow \text{TOPSIS}(G, k, w, K)$
2. Use each influential nodes  $v \in P$  as a centroid of a community  
 $CP_{centroid} \leftarrow \{\{v\} | v \in P\}$
3. Enlarge communities into  $CP_{centroid}$  :  
 $CP \leftarrow \text{Enlarging}(G, CP_{centroid})$
4. Return communities detected  $CP$

**End**

The main two steps in the proposed method are influential nodes detection using TOPSIS and then the community detection using k-means. These two steps are respectively highlighted in Algorithm 2 and Algorithm 3.

To get nodes' influence ranking using TOPSIS methodology involves four centrality measures as multi-attribute criteria. These measures are degree centrality (DC), betweenness centrality (BC), closeness centrality (CC) and eigen-vector centrality (EC) that have already been introduced in the subsection III.B. Algorithm 3 details technically the five steps highlighted in the subsection III.A.

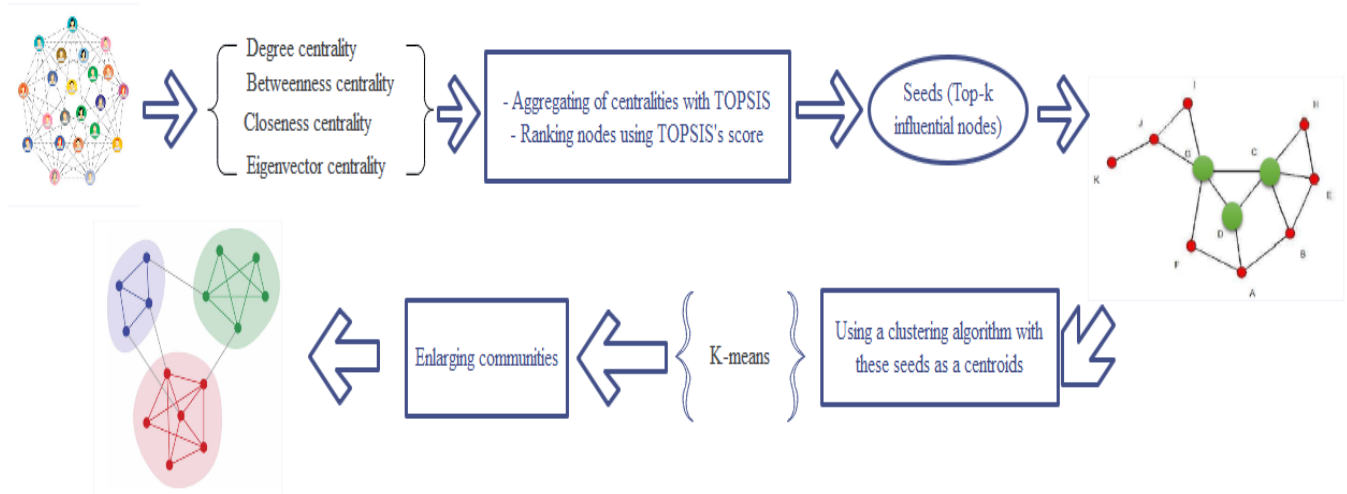


Fig. 4. The research methodology for seeds selection and expansion

**Algorithm 3:** Selection of top-K influential nodes as centroids using TOPSIS' scores.

**Input:**  $G(V, E)$  : A social network,  
 $K$  : the number of influential nodes (number of communities),  
 $k = \{k_1, k_2, \dots, k_n\}$  : a set of centrality measures,  
 $w = \{w_1, w_2, \dots, w_n\}$  : a set of weights assigned to centralities.

**Output:**  $P$  top-K influential nodes (a set of centroids)  
**Begin**

1. Create a decision matrix  $D$  based on  $k$  measures of centrality where  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, k$  and  $n$  is the number of nodes

$$D = \begin{bmatrix} d_{11} & \dots & d_{1k} \\ \vdots & \ddots & \vdots \\ d_{n1} & \dots & d_{nk} \end{bmatrix}$$

2. Normalize the decision matrix  $D$ :

$$\text{For each } d_{ij} \text{ in } D \text{ do } t_{ij} = \frac{d_{ij}}{\sqrt{\sum_{i=1}^k d_{ij}^2}}$$

$$N = \begin{bmatrix} t_{11} & \dots & t_{1k} \\ \vdots & \ddots & \vdots \\ t_{n1} & \dots & t_{nk} \end{bmatrix}$$

3. Make the decision matrix normalized and weighted to obtain the matrix  $R$

$$\text{For each } t_{ij} \text{ in } N \text{ do } r_{ij} = t_{ij} * w_j$$

$$R = \begin{bmatrix} r_{11} & \dots & r_{1k} \\ \vdots & \ddots & \vdots \\ r_{n1} & \dots & r_{nk} \end{bmatrix}$$

4. Calculate the ideal best solution  $A^+$  and negative ideal solution  $A^-$  (centrality measures are taken as benefit attributes)

$$A^+ = \max(r_{ij}) \text{ and } A^- = \min(r_{ij})$$

5. Calculate separation from ideal and negative solutions

$$Sep_p = \sqrt{\sum_{j=1}^k (r_{ij} - A^+)^2} \text{ and}$$

$$Sep_n = \sqrt{\sum_{j=1}^k (r_{ij} - A^-)^2}$$

6. Calculate the relative closeness score of each node

$$\text{For each } r_{ij} \text{ in } R \text{ do } C_i = \frac{Sep_n}{Sep_n + Sep_p}$$

7. Rank nodes based on  $C_i$  the higher score means important node
8. Select  $K$  nodes using top- $K$   $C_i$  :  $P \leftarrow top\_k(V, C_i, K)$

**End**

The output of this algorithm is the first  $k$  influential nodes. A node's influence depends on its score. The higher the score is, the more influential the node is. The idea of the proposed approach is to use the top- $K$  influential nodes as initial centroids to start  $k$ -means algorithm in order to detect communities in the network. The number of these nodes can be specified by the user or found automatically using the elbow method or silhouette method.

The process of community detection in the proposed approach is given in Algorithm 3. Communities within a network can be modeled as groups of individuals that have a kind of similarity. In this paper this similarity is computed by the Euclidean distance between these individuals. The value given to  $k$  will correspond to the number of the output communities. Each top- $K$  influential node resulted from Algorithm 2 is considered as a reference to all the nodes of the network. Euclidean distances are computed between each centroid and all the other nodes. Each centroid gathers around it the nodes that are closest to it, and the new centroid of each cluster is computed by the mean of all its elements. These steps are repeated until there are no new computed centroids. The last clusters are the detected communities in the network and their centroids are the resulted seeds.

**Algorithm 4:** Computing communities using k-means

**Input:**  $(V, E)$  : A social network,  
 $CP_{centroid}$  : Top-K influential nodes (the output of Algorithm 2)

**Output:**  $CP$  detected communities

**Begin**

1. Initialize the centroids with Top-k influential nodes  
/\*number of communities to be found/
2. For each community  $CP_j$ , REPEAT
  - 2.1 Assign each node  $v$  to the communities which has the closest mean based on their centrality measures

$$c^{(i)} = \arg \min_j \|v^{(i)} - CP_{centroid}^i\|^2$$

- 2.2 Compute new centroids for each community

$$CP_{centroid}^i = \frac{\sum_{i=1}^m 1\{c_{(i)}=j\}v^i}{\sum_{i=1}^m 1\{c_{(i)}=j\}}$$

3. UNTIL convergence criteria is reached
4. END FOR
5. Return detected communities  $CP$

**End**

V. EXPERIMENTAL SETUP

In this section a series of experiments are managed to show the efficiency of the proposed approach. In the beginning network with ground truth is used in order to validate the proposed approach referring to the reference of each network. Thereafter, this approach is applied to a larger network without ground truth to find different communities within it. These series of experiments have been performed using Python (3.7.14) as a tool of implementation. The use of Python is argued by its richness by a large number of useful libraries that make the data analysis and computing with visualization easier and simplest. For instance, in these experimentations we took advantage of the libraries; NetworkX for the manipulation of the complex network, Numpy for the scientific computing, Pandas for data analysis, Matplotlib for visualizations, Scikit-learn for clustering, CDlib for evaluating communities.

A. Networks Presentation

1) *Facebook ego network*: This dataset consists of 'circles' (or 'friends' lists) from Facebook. It includes node features (profiles), circles, and ego networks. It's downloadable from the Stanford large network dataset collection [42]. Fig. 5 presents this dataset. It has 4039 nodes and 88234 edges. Table II presents some characteristics of the dataset.

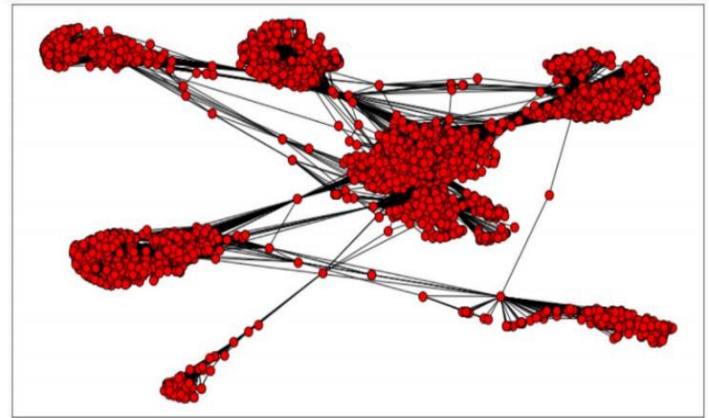


Fig. 5. Facebook ego network

TABLE II. CHARACTERISTICS OF THE DATASET

NODES	4039
EDGES	88234
NODES IN LARGEST WCC	4039 (1.000)
EDGES IN LARGEST WCC	88234 (1.000)
NODES IN LARGEST SCC	4039 (1.000)
EDGES IN LARGEST SCC	88234 (1.000)
AVERAGE CLUSTERING COEFFICIENT	0.6055
NUMBER OF TRIANGLES	1612010
FRACTION OF CLOSED TRIANGLES	0.2647
DIAMETER (LONGEST SHORTEST PATH)	8
90-PERCENTILE EFFECTIVE DIAMETER	4.7

2) *Zachary karate club network*: It's a real-world network that is well-known in the domain of community structure [43]. The data was collected from a university of karate club in 1977 by Zachary. It represents relationships between members of the karate club. The network is split into two groups after a dispute between the two masters John A and Mr. Hi. Table III presents the characteristics of this dataset.

TABLE III. CHARACTERISTICS OF ZACHARY DATASET

Nodes  N	Edges	Max degree	Average degree	Diameter
34	78	17	4.588 24	5

### B. Results and Discussion

This subsection details all the steps of the proposed approach with the two versions of k-means applied on the datasets presented above. We start the experimentation on Facebook dataset and then we validate it on Zachary network that has a ground truth of the detected communities within it.

1) *Top-K influential nodes detection*: Facebook data has been anonymized by replacing the Facebook-internal ids for each user with a new value. For each ego-network, files for circles, edges, egofeat, feat and featnames are provided. As we are going to cluster people only by their friendship, we only consider the edges file. The first step to implement TOPSIS is the construction of the evaluation matrix. Centrality measures are calculated beforehand and concatenated in a decision matrix. The measures DC, BC, CC and EC constitute the columns of this matrix, and each node in the network constitutes one of its rows. This matrix will be normalized and weighted in the next step to be ready for the following steps, see Algorithm 3.

After implementing and running TOPSIS methodology, we consider  $k=10$  to return the top-10 influential nodes presented by descending sort in Table IV. These nodes are numbered in the column named "Node" and their scores are given in the column named "TOPSIS". The ranking of these nodes is as following: 107>1684>1912>3437>0>1085>698>567>58>428, with ">" means more influential.

TABLE IV. TOP-10 INFLUENTIAL NODES USING TOPSIS

Node	SCORE
107	0,913277
1684	0,695566
1912	0,496063
3437	0,488865
0	0,304321
1085	0,297379
698	0,231106
567	0,193646
58	0,169464

428	0,132923
-----	----------

The Susceptible-Infected (SI) model is used to look at the spreading effect of top-K influential nodes in order to assess the effectiveness of the ranking model. The SI model is frequently used to study the dynamics of epidemics on networks. Every node in the SI model has two distinctive states:

a) Susceptible  $S(t)$  indicates the number of people who are susceptible to the disease but have not yet obtained it;

b) Infected  $I(t)$  reflects the number of people who have contracted the disease and are able to disseminate it to susceptible people. For each contaminated node, one randomly sensitive neighbor contracts the disease with probability at each step (here,  $\lambda = 0.3$  for uniformity).

For this epidemic model,  $\lambda$  indicates the range across which a node can have an effect on epidemic spreading on networks. Through the intermediaries, an infected node can spread the infection not just to its immediate neighbors but also to its higher order neighbors. In this mode,  $F(t)$  stands for the number of contaminated nodes at time  $t$ . Using different initially infected nodes, the number of infected nodes should be equal to the overall number of nodes in networks. The average number of infected nodes at each iteration or the spreading rate is the indicator to assess the influence of the initial infected node. The proposed technique is compared with degree, closeness, betweenness, and eigenvector centrality using the SI model. Each implementation identifies the top 10 nodes to infect, and then the SI model is used to determine how the information spreads throughout the network. The influence of the nodes that either presents in the top-10 rankings by the proposed model and the four metrics of centrality are studied. In order to increase the precision of the results, the algorithm will be repeated 10000 times. Thus, the variation (standard deviation) and the mean of each iteration will be calculated. Fig. 6 presents the results.  $F(t)$  is the cumulative infected nodes The simulations are on  $F(t)$  as a function of time for the proposed network.  $F(t)$  increases with time and finally reaches the steady value. According to Fig. 6, the proposed method outperforms DC, CC and EC. The results between BC and the proposed approach are close. Their lines almost overlap as shown in the Fig. 6 and the members of their top-10 lists are the same.

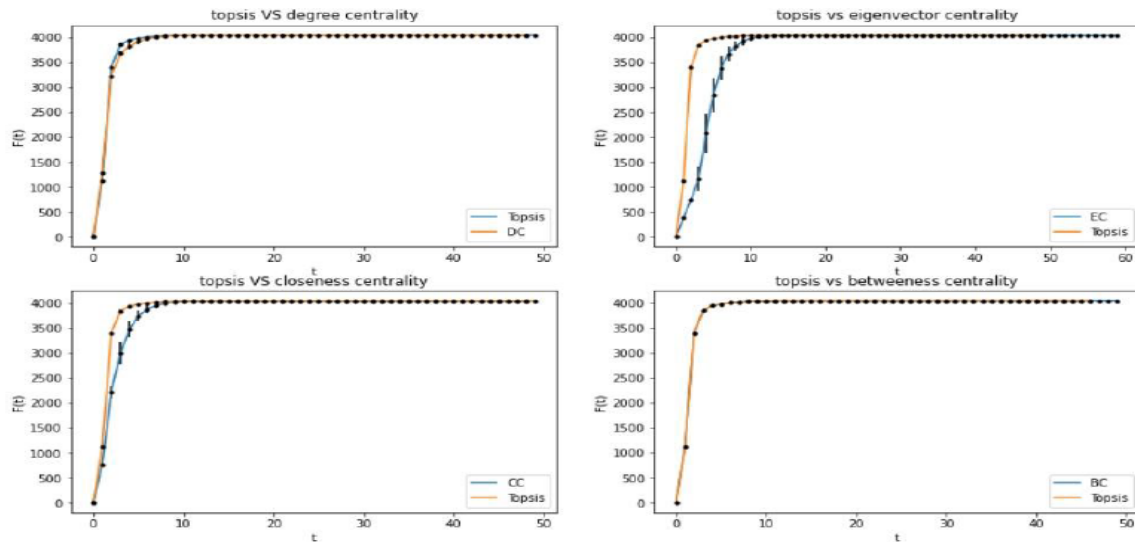


Fig. 6. The cumulative number of infected nodes as a time function, with the initially infected are the top-10 list by the proposed method, DC, CC, BC, and EC. Results are obtained by averaging over 100000 implements ( $\beta=0.3$ ).

2) *Community detection using static k-means*: The second step of the proposed approach is to run k-means algorithm, with  $k=10$ . The starting centroids for the k-means are the top-10 influential nodes computed by TOPSIS methodology and given in Table IV. Concretely, these nodes are:

$$CP_{centroid}^0 = node_{107}, CP_{centroid}^1 = node_{1684}, \dots, CP_{centroid}^9 = node_{428}.$$

The k-means algorithm is applied in two ways; one with the centrality measures and the other with the k-means algorithm is applied in two ways; one with the centrality measures and the other with the adjacency matrix.

a) *K-means with centrality measures*: In this first way of application of k-means, the centrality measures of top-10 influential nodes are calculated as a training data. Fig. 7 demonstrates the clustering using centrality measures, the big nodes with different colors present the top-10 influential nodes.

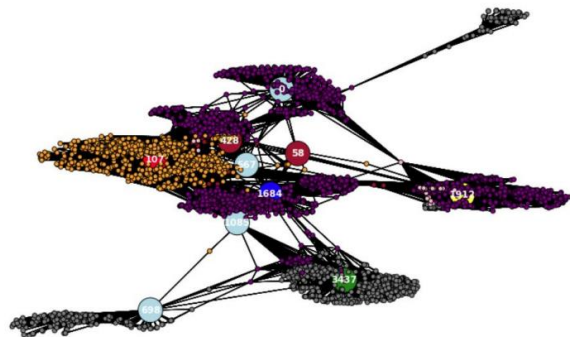


Fig. 7. Clustering based on k-means using centrality measures with initialization with top-10 influential nodes

b) *K-means with adjacency matrix*: We're going straight in this step to construct the adjacency matrix of the network, then the matrix of the top-10 nodes. An adjacency matrix is a way of representing a graph as a matrix of Booleans; 0 when two nodes are not connected, and 1 when two nodes are connected.

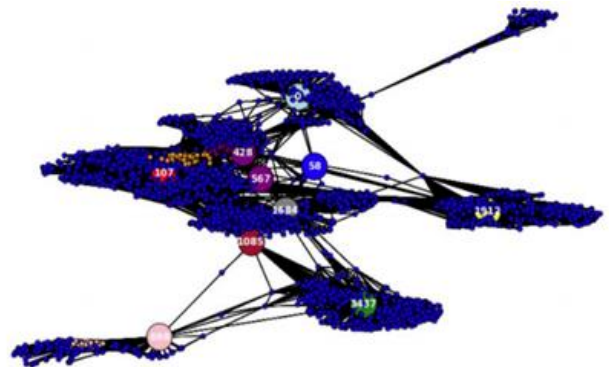
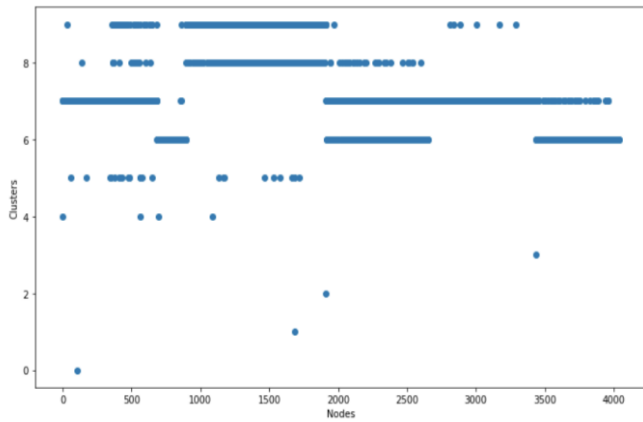
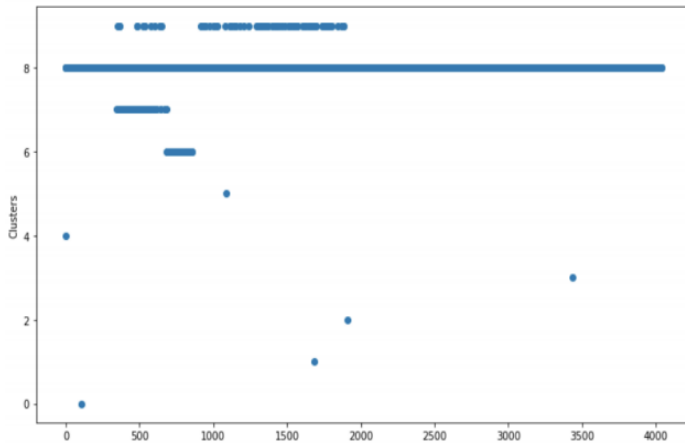


Fig. 8. Clustering based on k-means using adjacency matrix with initialization with top-K influential nodes

According to Fig. 8, it appears that implementing k-means using centrality measures gives good results in terms of distribution of nodes than using adjacency matrix. Although the algorithm is initialized with Top-10 influential nodes, one dominant cluster (blue cluster) is obtained which explains that the majority of nodes are grouped in this cluster. To confirm this result we resort to the scatterplot of these two approaches. The scatterplot below in Fig. 9(a) shows the distribution of nodes by their cluster.



(a)The scatterplot of nodes and their cluster with initialization of top-10 influential nodes using centrality measure



(b)The scatterplot of nodes and their cluster with initialization of top-10 influential nodes using adjacency matrix

Fig. 9. Distribution of nodes in their clusters using centrality measures and adjacency matrix

From Fig. 9(b) the cluster 8 contains the majority of the network’s nodes; the other nodes are distributed to the cluster 9, the cluster 7 and the cluster 6. Other clusters contain just one node. Fig. 9(a) demonstrates the distribution of nodes using centrality measures as node features for k-means training data. Nodes in this clustering are not also well distributed and the top-10 nodes have been isolated in 6 separate clusters as Table V shows.

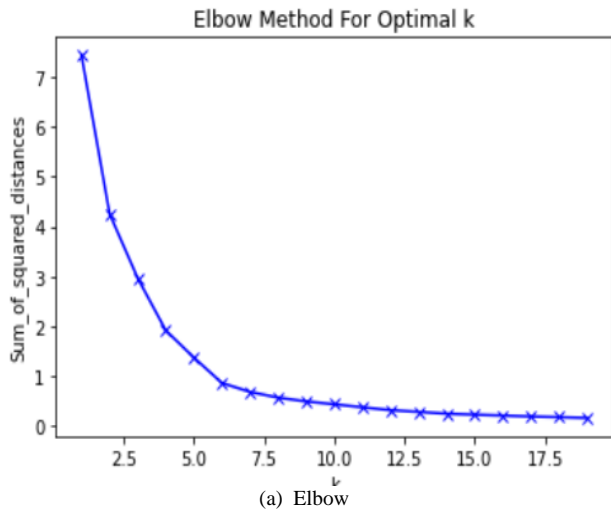
In this clustering some of the top-10 nodes are the lonely nodes in their cluster like 107, 1684, 1912 and 3437. While the nodes 0, 1085, 567 and 698 are gathered in one cluster and the two other nodes of the top-10 nodes are grouped with some few nodes in the same cluster. The k-means algorithm using the centrality measures clusters the nodes based on their influence in the network, that’s why we notice that the more

the nodes have more influence in the network, the more they are isolated in other clusters. This kind of clustering can be applied to minimize the influence in social networks. The influence minimization can predict the spread of deprecatory rumors, fake news, and spread of disease.

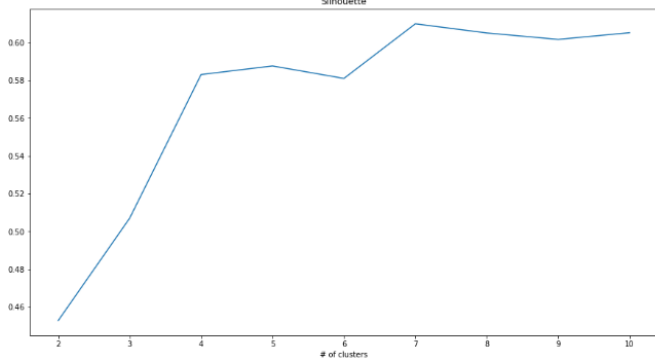
TABLE V. DISTRIBUTIONS OF INFLUENTIAL NODES IN THEIR FINAL CLUSTERS

Node	TOPSIS	Cluster
107	0.913277	0
1684	0.695566	1
1912	0.496063	2
3437	0.488865	3
0	0.304321	4
1085	0.297379	4
698	0.231106	4
567	0.193646	4
58	0.169464	5
428	0.132923	5

3) *Community detection using dynamic k-means:* Although the proposed approach is simple and fast, yet there are also some limitations. From the scatterplots in Fig. 9, we conclude that the value 10 given to k is big and not precise. We need then to assign an optimal value to k before starting the k-means algorithm. It’s hard to know this value from the beginning, reason for what we need to apply the dynamic K-means. As discussed in the section D, we implement Elbow and Silhouette methods for this purpose. The two approaches based on the centrality measures and also the adjacency matrix is also used in this context. Centrality measures exceeds adjacency matrix in terms of defining the optimal-k. The problem of working with the adjacency matrix is the instability of defining the optimal k because we get different values of k for each execution. Hence, we focus on centrality measures to define the optimal value for k. The value generated using the Elbow method and Silhouette method are respectively k=6 and k=7, Fig. 10. From the previous experiments where k=10, we concluded that big values of k don’t give good clustering. So we consider the optimal value of k is 6 and then we apply the proposed approach using the first six influential nodes as starting centroids for k-means algorithm. The output of the proposed approach into six clusters is highlighted in Fig. 11 below.



(a) Elbow



(b) Silhouette

Fig. 10. The clustering of Facebook ego network using dynamic k-means

According to the graph in Fig. 11, it seems that we have four dominant communities in the network which is confirmed by the scatterplot in Fig. 12. The distribution of nodes to the detected communities looks reasonable in this results compared to the previous one. This means that when  $k$  is well defined, the proposed approach gives better results. These results need to be confirmed on datasets with ground truth.

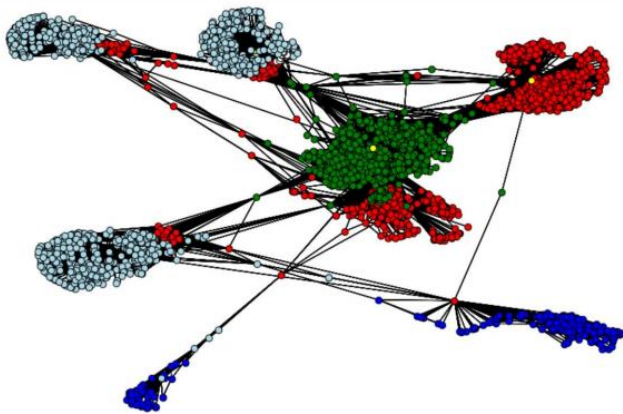


Fig. 11. The clustering of facebook ego network using dynamic k-means

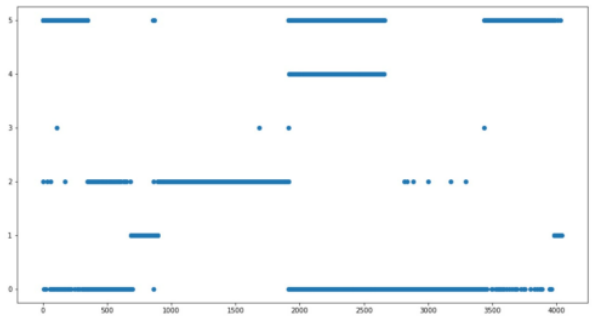


Fig. 12. The scatterplot of Facebook ego network using dynamic k-means

4) *Evaluation metrics:* It is necessary to use specific performance measures to measure the similarity between the two dataset partitions. In this experimental study, two classical measures are used.

The first one is the Rand index [44]. It is the portion of point pairs  $(x_1, x_2)$  that are organized similarly in both divisions. Either  $x_1$  and  $x_2$  are members of the same cluster in both the situations or they are members of different clusters.

The Rand index can be inflated artificially by predicting many clusters. Numerous pairs of points will belong to separate clusters, and the possibility that two points with different labels will be found in two different clusters will be considerable. This effect is remedied by The Adjusted Rand Index (*ARI*), which normalizes the Rand Index (*RI*) [45]:

$$ARI = \frac{RI - E(RI)}{\max(RI) - E(RI)} \quad (5)$$

Where  $E(RI)$  is the expectation of the value of the Rand index, in other words, the index obtained by randomly splitting the data. This adjusted index is only close to 1 when the clustering exactly matches the original partition and is close to 0 otherwise.

The second measure used in this experimental study is the normalized mutual information [46]. In probability theory and information theory, the mutual information of two random variables is a quantity measuring of the statistical dependence of these variables. Normalized mutual information *NMI* is a variant of mutual information. Its value is between 0 and 1. The closer *NMI* is to 1, the closer the result is to the ground truth. It can be defined as:

$$NMI(X, Y) = \frac{I(X, Y)}{\sqrt{H(X)H(Y)}} \quad (6)$$

Where,  $I()$  is the mutual information and  $H()$  is the entropy.

5) *Experiments on dataset with ground-truth:* The evaluation of detected communities is still an open issue in the scientific community because of the lack of models and references. Since the availability of a ground truth community structure for large real networks is difficult, we choose to validate the proposed approach on Zachary karate club network.

In this experiment, we rely on the dynamic k-means to get the optimal value of  $k$ . In this case the optimal  $k=2$  by the silhouette method as shown in Fig. 13. We test k-means with top-2 influential nodes using centrality measures but the



results are not satisfactory reason for which we use dynamic k-means with the adjacency matrix on Zachary network. Fig. 15 shows the obtained results and Fig. 14 gives the ground-truth community structure of Zachary karate club.

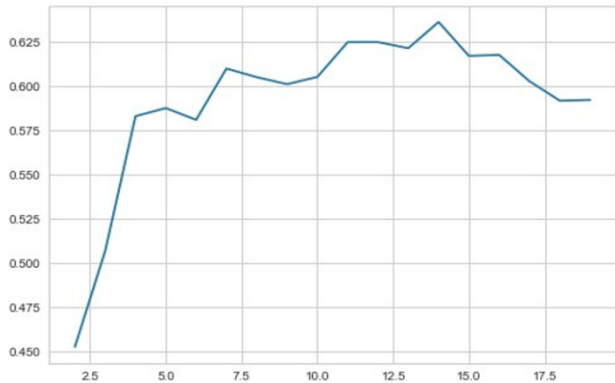


Fig. 13. The optimal-k of Zachary dataset given by silhouette method

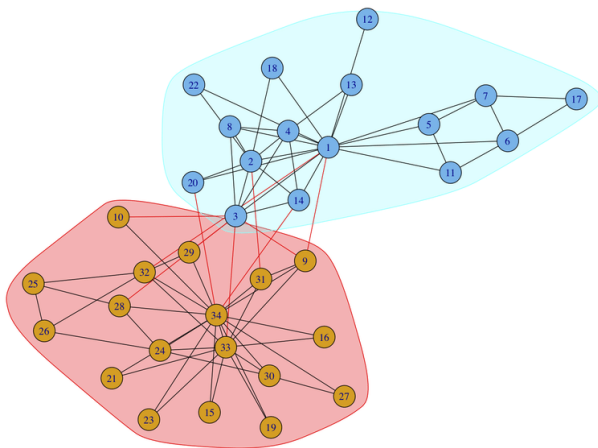


Fig. 14. The ground-truth community structure of Zachary karate club

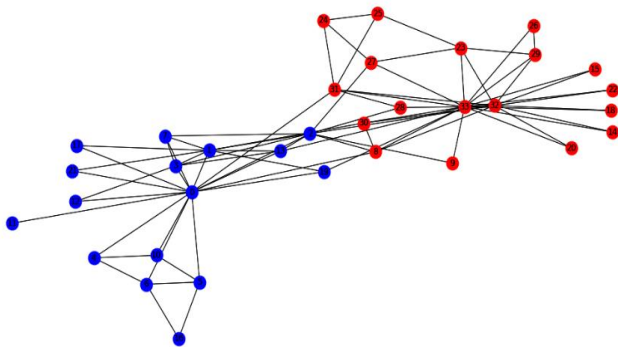


Fig. 15. The community structure detected by the proposed approach

Table VI presents the obtained results on Zachary compared to other algorithms of community detection like Louvain [47], Newman [48], Walktrap [5], Licod [18] and Yasca [27]. Compared to the ground-truth community structure, the proposed approach also detects 2 communities in

Zachary as displayed in Fig. 15. Although some nodes deviate from the ground-truth community structure, the quality of the discovered communities is higher than other algorithms as mentioned in Table VI. NMI and ARI values are higher compared to other community detection algorithms.

TABLE VI. COMPARISON BETWEEN ALGORITHMS OF COMMUNITY DETECTION AND THE PROPOSED APPROACH ON ZACHARY KARATE CLUB.

Zachary (2 groups)	Algorithm	ARI	NMI	Detected communities
	Newman	0,46	0,57	5
	Louvain	0,46	0,58	4
	Walktrap	0,33	0,50	5
	Licod	0,62	0,60	3
	YASCA	0,69	0,77	2
	<b>Proposed ML approach</b>	<b>0,88</b>	<b>0,83</b>	2

## VI. CONCLUSION

In this paper, a new seed-expanding algorithm to detect communities in social networks is introduced. The selection of seeds is based on centrality measures, gathered as multi-attribute in TOPSIS. The Top-K influential nodes obtained by TOPSIS methodology are used as starting centroids to run the k-means algorithm. In the experimental study, k-means is implemented using centrality measures and adjacency matrix as training data to compare the results. We notice that on large real networks like Facebook, the centrality measures behave well than adjacency matrix in terms of nodes distribution except that we found difficulties to determine the good values of k for the clustering. That is why we proceed to dynamic k-means using Elbow and silhouette methods to get the optimal values of k. Thereafter we observe that the results seem more reasonable with the dynamic k-means. Eventually, for assessing the quality of the discovered communities, we use Zachary karate club to validate the proposed model instead of Facebook because of the unavailability of a ground-truth for Facebook. ARI and NMI are two measures used for the test, and the results show that the proposed approach is better than other available algorithms for community detection. Many possible future directions have been opened up by this work. For example, implementing other clustering methods for the purpose of detecting communities such as K-medoids, C-means, fuzzy k-means and compare the results with the K-mean algorithms. As we can extend this work on other networks apart from social networks by using other centralities in the initial phase which are adequate for these types of networks.

## REFERENCES

- [1] D. J. Watts et S. H. Strogatz, « Collective dynamics of “small-world” networks », *Nature*, vol. 393, no 6684, p. 440-442, juin 1998, doi: 10.1038/30918.
- [2] A.-L. Barabási et R. Albert, « Emergence of Scaling in Random Networks », *Science*, vol. 286, no 5439, p. 509-512, oct. 1999, doi: 10.1126/science.286.5439.509.

- [3] S. Bilke et C. Peterson, « Topological properties of citation and metabolic networks », *Phys. Rev. E*, vol. 64, no 3, p. 036106, août 2001, doi: 10.1103/PhysRevE.64.036106.
- [4] J. Eriksen, M. K. Jensen, P. Sjøgren, O. Ekholm, et N. K. Rasmussen, « Epidemiology of chronic non-malignant pain in Denmark », *Pain*, vol. 106, no 3, p. 221-228, déc. 2003, doi: 10.1016/S0304-3959(03)00225-2.
- [5] P. Pons et M. Latapy, « Computing communities in large networks using random walks », p. 20.
- [6] G. Cordasco et L. Gargano, « Label propagation algorithm: a semi-synchronous approach », *Int. J. Soc. Netw. Min.*, vol. 1, no 1, p. 3, 2012, doi: 10.1504/IJSNM.2012.045103.
- [7] M. E. J. Newman et M. Girvan, « Finding and evaluating community structure in networks », *Phys. Rev. E*, vol. 69, no 2, p. 026113, févr. 2004, doi: 10.1103/PhysRevE.69.026113.
- [8] S.-J. Chen et C.-L. Hwang, « Fuzzy Multiple Attribute Decision Making Methods », in *Fuzzy Multiple Attribute Decision Making*, vol. 375, Berlin, Heidelberg: Springer Berlin Heidelberg, 1992, p. 289-486. doi: 10.1007/978-3-642-46768-4\_5.
- [9] L. C. Freeman, « Centrality in social networks conceptual clarification », *Soc. Netw.*, vol. 1, no 3, p. 215-239, janv. 1978, doi: 10.1016/0378-8733(78)90021-7.
- [10] Ulrik Brandes, « A faster algorithm for betweenness centrality », *The Journal of Mathematical Sociology*, p. 163-177, 2001.
- [11] P. Bonacich et P. Lloyd, « Eigenvector-like measures of centrality for asymmetric relations », *Soc. Netw.*, vol. 23, no 3, p. 191-201, juill. 2001, doi: 10.1016/S0378-8733(01)00038-7.
- [12] R. Nainggolan, R. Perangin-angin, E. Simarmata, et A. F. Tarigan, « Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method », *J. Phys. Conf. Ser.*, vol. 1361, no 1, p. 012015, nov. 2019, doi: 10.1088/1742-6596/1361/1/012015.
- [13] P. J. Rousseeuw, « Silhouettes: A graphical aid to the interpretation and validation of cluster analysis », *J. Comput. Appl. Math.*, vol. 20, p. 53-65, nov. 1987, doi: 10.1016/0377-0427(87)90125-7.
- [14] H. Gupta et M. K. Barua, « Supplier selection among SMEs on the basis of their green innovation ability using BWM and fuzzy TOPSIS », *J. Clean. Prod.*, vol. 152, p. 242-258, mai 2017, doi: 10.1016/j.jclepro.2017.03.125.
- [15] « (26) Analizing Topsis Method for Selecting the Best Wood Type | Ria Sari - Academia.edu ». [https://www.academia.edu/38541712/Analizing\\_Topsis\\_Method\\_for\\_Selecting\\_the\\_Best\\_Wood\\_Type](https://www.academia.edu/38541712/Analizing_Topsis_Method_for_Selecting_the_Best_Wood_Type) (consulté le 21 avril 2022).
- [16] F. El Allaki, J. Christensen, et A. Vallières, « A modified TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) applied to choosing appropriate selection methods in ongoing surveillance for Avian Influenza in Canada », *Prev. Vet. Med.*, vol. 165, p. 36-43, avr. 2019, doi: 10.1016/j.prevetmed.2019.02.006.
- [17] A. Kelemenis et D. Askounis, « A new TOPSIS-based multi-criteria approach to personnel selection », *Expert Syst. Appl.*, vol. 37, no 7, p. 4999-5008, juill. 2010, doi: 10.1016/j.eswa.2009.12.013.
- [18] R. Kanawati, LICOD: Leaders Identification for Community Detection in Complex Networks. 2011, p. 582. doi: 10.1109/PASSAT/SocialCom.2011.206.
- [19] S. Papadopoulos, Y. Kompatsiaris, et A. Vakali, « A Graph-Based Clustering Scheme for Identifying Related Tags in Folksonomies », in *Data Warehousing and Knowledge Discovery*, Berlin, Heidelberg, 2010, p. 65-76. doi: 10.1007/978-3-642-15105-7\_6.
- [20] B. Bollobas et O. Riordan, « Clique percolation », *Random Struct. Algorithms*, vol. 35, no 3, p. 294-322, oct. 2009, doi: 10.1002/rsa.20270.
- [21] D. Shah et T. Zaman, « Community Detection in Networks: The Leader-Follower Algorithm », *ArXiv*, 2010.
- [22] D. Parthasarathy, D. Shah, et T. Zaman, « Leaders, Followers, and Community Detection », p. 9, 2018.
- [23] M. Danisch, J.-L. Guillaume, et B. Le Grand, « Unfolding Ego-Centered Community Structures with "A Similarity Approach" », in *Complex Networks IV*, vol. 476, G. Ghoshal, J. Poncele-Casasnovas, et R. Tolksdorf, Éd. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, p. 145-153. doi: 10.1007/978-3-642-36844-8\_14.
- [24] J. J. Whang, D. F. Gleich, et I. S. Dhillon, « Overlapping community detection using seed set expansion », in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management - CIKM '13*, San Francisco, California, USA, 2013, p. 2099-2108. doi: 10.1145/2505515.2505535.
- [25] M. Weskida et R. Michalski, « Evolutionary algorithm for seed selection in social influence process », in *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Davis, California, août 2016, p. 1189-1196.
- [26] Y. Wang, B. Zhang, A. V. Vasilakos, et J. Ma, « PRDiscount: A Heuristic Scheme of Initial Seeds Selection for Diffusion Maximization in Social Networks », in *Intelligent Computing Theory*, Cham, 2014, p. 149-161. doi: 10.1007/978-3-319-09333-8\_17.
- [27] R. Kanawati, YASCA: An Ensemble-Based Approach for Community Detection in Complex Networks. 2014. doi: 10.1007/978-3-319-08783-2\_57.
- [28] A. Zakrzewska et D. A. Bader, « A Dynamic Algorithm for Local Community Detection in Graphs », in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, New York, NY, USA, août 2015, p. 559-564. doi: 10.1145/2808797.2809375.
- [29] B. R. V. E. Kanaga, et P. Bródka, « Overlapping community detection using superior seed set selection in social networks », 10 août 2018.
- [30] S. Papadopoulos, Y. Kompatsiaris, A. Vakali, et P. Spyridonos, « Community detection in Social Media », *Data Min. Knowl. Discov.*, vol. 24, no 3, p. 515-554, mai 2012, doi: 10.1007/s10618-011-0224-z.
- [31] B. Bollobas et O. Riordan, « Clique percolation », *Random Struct. Algorithms*, vol. 35, no 3, p. 294-322, oct. 2009, doi: 10.1002/rsa.20270.
- [32] J. J. Whang, D. F. Gleich, et I. S. Dhillon, « Overlapping Community Detection Using Neighborhood-Inflated Seed Expansion », *IEEE Trans. Knowl. Data Eng.*, vol. 28, no 5, p. 1272-1284, mai 2016, doi: 10.1109/TKDE.2016.2518687.
- [33] C.-T. Chen, C.-T. Lin, et S.-F. Huang, « A fuzzy approach for supplier evaluation and selection in supply chain management », *Int. J. Prod. Econ.*, vol. 102, no 2, p. 289-301, 2006.
- [34] « Plant location selection based on fuzzy TOPSIS | SpringerLink ». <https://link.springer.com/article/10.1007/s00170-004-2436-5> (consulté le 16 mai 2022).
- [35] Y.-M. Wang et T. M. S. Elhag, « Fuzzy TOPSIS method based on alpha level sets with an application to bridge risk assessment », *Expert Syst. Appl.*, vol. 31, no 2, p. 309-319, août 2006, doi: 10.1016/j.eswa.2005.09.040.
- [36] T. Kaya et C. Kahraman, « Multicriteria decision making in energy planning using a modified fuzzy TOPSIS methodology », *Expert Syst. Appl. Int. J.*, vol. 38, no 6, p. 6577-6585, juin 2011, doi: 10.1016/j.eswa.2010.11.081.
- [37] S. Wasserman et K. Faust, *Social network analysis: methods and applications*. 1994.
- [38] J. Scott, *Social Network Analysis: A Handbook*. SAGE, 2000.
- [39] J. M. Kleinberg, « Authoritative sources in a hyperlinked environment », *J. ACM*, vol. 46, no 5, p. 604-632, sept. 1999, doi: 10.1145/324133.324140.
- [40] S. P. Borgatti et M. G. Everett, « Models of core/periphery structures », *Soc. Netw.*, vol. 21, no 4, p. 375-395, oct. 2000, doi: 10.1016/S0378-8733(99)00019-2.
- [41] N. Krislock et H. Wolkowicz, « Euclidean distance matrices and applications », in *Handbook on semidefinite, conic and polynomial optimization*, Springer, 2012, p. 879-914.
- [42] « SNAP: Network datasets: Social circles ». <https://snap.stanford.edu/data/ego-Facebook.html> (consulté le 17 mai 2022).
- [43] W. W. Zachary, « An Information Flow Model for Conflict and Fission in Small Groups », *J. Anthropol. Res.*, vol. 33, no 4, p. 452-473, 1977.
- [44] « Objective Criteria for the Evaluation of Clustering Methods on JSTOR ». <https://www.jstor.org/stable/2284239> (consulté le 13 juillet 2022).
- [45] L. Hubert et P. Arabie, « Comparing partitions », *J. Classif.*, vol. 2, no 1, p. 193-218, déc. 1985, doi: 10.1007/BF01908075.

- [46] « [PDF] Robust data clustering | Semantic Scholar ». <https://www.semanticscholar.org/paper/Robust-data-clustering-Fred-Jain/9e7a86fd9e15bf37d937a79ccb7efb78bb070f74> (consulté le 13 juillet 2022).
- [47] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, et E. Lefebvre, « Fast unfolding of communities in large networks », *J. Stat. Mech. Theory Exp.*, vol. 2008, no 10, p. P10008, oct. 2008, doi: 10.1088/1742-5468/2008/10/P10008.
- [48] M. E. J. Newman, « Fast algorithm for detecting community structure in networks », *Phys. Rev. E*, vol. 69, no 6, p. 066133, juin 2004, doi: 10.1103/PhysRevE.69.066133.

# Implementation of ICT Continuity Plan (ICTCP) in the Higher Education Institutions (HEI'S): SUC'S Awareness and its Status

Chester L. Cofino<sup>1</sup>, Ken M. Balogo<sup>2</sup>, Jeffrey G. Alegia<sup>3</sup>, Michael Marvin P. Cruz<sup>4</sup>,  
Benjamin B. Alejandro Jr.<sup>5</sup>, Felicisimo V. Wenceslao, Jr.<sup>6</sup>

College of Computer Studies, Central Philippines State University, Kabankalan City, Philippines<sup>1,2,3,4</sup>  
College of Industrial Technology, Negros Oriental State University, Dumaguete City, Philippines<sup>5</sup>  
College of Information and Computing Studies, Northern Iloilo State University, Iloilo, Philippines<sup>6</sup>

**Abstract**—The purpose of this study was to assess the level of awareness of the management and the personnel within the academic institution and identify the implementation status of the ICTCP to the implementing SUCs. The researchers used the BCM Framework was utilized in this study as the model for identifying the level of awareness of the personnel within the institution about the ICTCP. The research respondents were the personnel employed in the different States, Universities, and Colleges (SUCs) within the province of Negros Occidental. The respondents were selected through random sampling, they were provided by a google form link to answer the survey questionnaire. A total of thirty-five (35) IT personnel were included in the study's sample size. It was found out that most SUCs have consistent ICT system uptime because they can continuously provide services; surprisingly, this is independent of an ICT business continuity plan. Most SUCs do not entirely implement their ICT business continuity plans. Lastly, it is recommended that SUCs can significantly enhance service delivery if ICT business continuity planning is taken seriously, adopted, and entirely carried out.

**Keywords**—Business continuity plan (BCP); information, communication; and technology continuity Plan (ICTCP); state universities and colleges (SUCs); business continuity management (BCM) framework

## I. INTRODUCTION

The importance of technology for information and communication (ICT) as component of enterprises worldwide has become crucial. It comprises communication technologies, infrastructure, hardware, and software related to information systems. This technology and systems will eventually malfunction due to unexpected catastrophes or causes. The ICT systems must be recovered and put back into operation with the fewest possible downtimes because they are a crucial component of corporate functions. According to [1], the speed at which business operations recover guarantees that it will outperform its rivals and provide customers with a high degree of satisfaction.

In the Philippine setting, one of the legal bases is the Data Privacy Act of 2012 (RA10173), ensuring that all government institutions, like States, Universities, and Colleges (SUCs), will craft its ICT Continuity Plan (ICTCP) to protect their business process and provide excellent service to their clients.

Furthermore, on March 24, 2020, Republic Act No. 11469, known as the "Bayanihan to Heal as One Act" was signed into law. In this regard, the DICT hereby directs all public telecommunication entities (PTEs) and government agencies, including academic institutions, to submit their Business Continuity Plan (BCP) and other measures to ensure uninterrupted service and to address the increased demand for ICT services.

In this regard, this study aimed to assess the level of awareness of the management and the personnel within the academic institution and identify the implementation status of the ICTCP to the implementing SUCs. Ascertain the extent of the use of computers using ICT at the State Universities and Colleges (SUCs) in Negros Occidental, Philippines. Ascertain how frequently the ICT systems SUCs relied on broke down and if the continuity plan is in place. Lastly, to ascertain how an ICT business continuity plan affected the provision of services.

## II. RELATED WORKS

ICT business continuity planning is a system of practices, guidelines, and proactive preparation that guarantees the restoration of vital ICT-dependent services in an emergency [2]. It enables the company to make crucial strategic, tactical, and operational decisions on the availability of essential systems [3]. In preparation for any system outage, resources such as software, hardware, technical experience, monitoring, time, and other infrastructure are established during an ICT business continuity plan.

The implementation must be maintained periodically and managed adequately. However, some studies concerning IT Audit found that some of the existing Business Continuity Plan (BCP) that is not entirely and correctly updated and will fail the business to continue its operation when disruption happens [4]. This was also seconded by the study [2] that due to the ICT business continuity plan not being viewed as a strategic component of the company, it has not been completely implemented and was neglected. According to [5], many organizations did not feel the need to deal with the BCP in the past because the dependence on ICT was not so significant, and production could run for some time regardless of a data network in the organization.

BCP was beneficial for all of the organizations that were surveyed. Still, more focus is required on managing societal and individual impacts, building employee resilience, identifying influential crisis leaders, right-sizing plans, and planning to take advantage of opportunities after a disaster. It is also essential to assess employees' awareness of the general coordination of information technology regarding these processes [6].

### III. IMPLEMENTATION

The study's general objective is to ascertain how ICT continuity plans affects on service delivery and assess the level of awareness of the personnel of State, Universities, and Colleges (SUCs) in the Philippines, specifically in Negros Occidental. Furthermore, it is to identify the advantages of an ICT business continuity strategy for SUCs in Negros Occidental. The study is based on the descriptive type, which involved a primary gathering of data from the respondents in terms of their profile and assessment of the availability of the ICTCP in their respective institutions.

#### A. The Business Continuity Plan (BCM) Framework

The BCM Framework was utilized in this study as the model for identifying the level of awareness of the personnel within the institution about the ICTCP (Fig. 1). The six elements of the BCM life cycle, according to the international standards, the management of the business continuity program, instilling knowledge and skills into the organizational culture, knowing the organization, choosing options for business continuity, creating and putting into practice a business continuity response, and practicing and testing the developed plans [7].

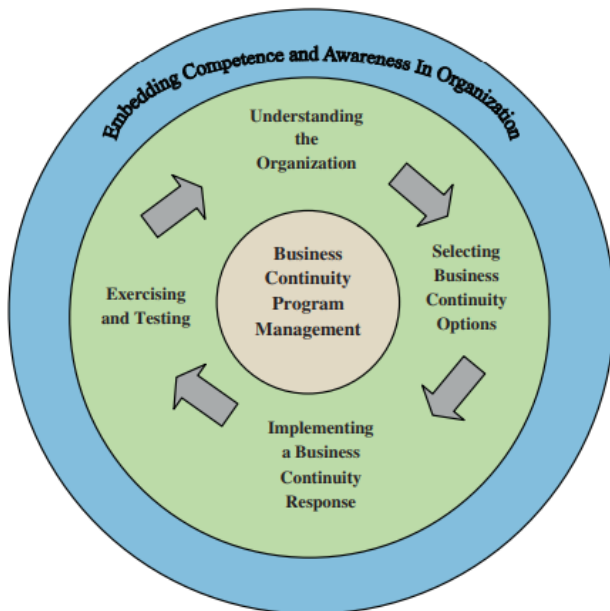


Fig. 1. The lifecycle of BCM [7].

#### B. Understanding the Organization

According to [8], the outcomes of the BCP implementation show that all framework components can be implemented, but some activity changes must be made to suit the organizational

conditions. Personnel in the organization must be aware of the implementation because they are the ones providing the services. This component aims to offer crucial details that will aid in understanding the company's goods, resources, and services [9]. Furthermore, the principal target objective, specific objectives, and policy are visible and disseminated to the employee from the top management to the office personnel.

Table I presents the role and responsibilities of the BCP team for SUCs. The team composed of the top management who see to it the requirements, budget allocation, and implementation of the ICTCP would be successful.

TABLE I. ROLE AND RESPONSIBILITIES FOR BCP TEAM

BCP Member	Roles
BCP Manager	<ul style="list-style-type: none"> <li>Ensures that the ICTCP is established, maintained, and reviewed periodically.</li> <li>Approves allocation of resources to ensure successful implementation of the plan.</li> </ul>
Vice President for Administration and Finance	<ul style="list-style-type: none"> <li>Ensures implementation and compliance with the ICTCP.</li> <li>Ensures the continual improvement of the ICTCP.</li> </ul>
Planning Officer	<ul style="list-style-type: none"> <li>Consolidates and evaluates personnel needs required to ensure the continuity of ICTCP critical unit functions and operations.</li> <li>Ensures continual improvement.</li> </ul>

Table II shows the composition of the ICT Response Team. The team must ensure that the ICTCP implementation is successful. The SUCs may designate a working team with expertise in ICT.

TABLE II. ROLE AND RESPONSIBILITIES OF ICT RESPONSE TEAM

Member	Roles
IT/MIS Officer	Establish, implement, and maintain a continuity of the ICT Business Operation of the University.
Data Privacy Officer	Inform and provide advice on data protection obligations, assist the University in monitoring internal compliance.
System Administrator	System administrators are the guardians of an organization and its data, ensuring that internal systems are safe and secure and are shielded from attacks and viruses.
Network Administrator	Ensuring ICT equipment remains updated and providing solutions to restore functionality.
Development Communication and Information Officer	Consolidates, controls, and validates all official information for dissemination.  Creating information, education, and communication (IEC) items in both English and regional dialects.

#### C. Selecting Business Continuity Options

Threats could appear out of nowhere at any time. Business continuity management is necessary to prepare for and address the problem. The last step in completing business continuity management is the requirement for a framework for a business continuity strategy. This framework will guide the organization using the business continuity plan document to

address threats or disasters [10]. Good decision on what framework to adapt is vital for the success of the implementation of the ICTCP.

#### D. Implementing a Business Continuity Response

According to [11], the fundamental idea behind business continuity is that to maintain company operations at an acceptable level, an organization must have the strategic and tactical capabilities to plan for and respond to business accidents and disruptions. Readiness is also vital in implementing a business continuity plan from management perspective and fulfilling the gap existing with a business continuity plan using a standard tool [12]. Many SUCs appear unfamiliar with and uncertain about which Business Continuity Management (BCM) option can be used for implementation. Organizations are becoming increasingly aware that being unprepared to handle disruptive occurrences could have disastrous results. BCM is a novel strategy to accomplish this goal. When a BCM is implemented in a company, there are three primary stages that need to be taken. The key items for the company should be recognized first. The Business Impact Analysis (BIA) technique can assist in this step by assisting in the systematic selection of those crucial products. Second, by carrying out the risk assessment process and creating the risk matrix, those risks that endanger the delivery of crucial products should be recognized and categorized. Finally, a BCP must be chosen for each disruption risk listed in the BCP section of the risk matrix and poses a danger to a significant product [13].

#### IV. RESULTS

The research respondents were the personnel employed in the different States, Universities, and Colleges (SUCs) within the province of Negros Occidental. The study utilized an adopted questionnaire from the study [2] on ICT Business Continuity Plan and Service delivery.

The respondents were selected through random sampling, they were provided by a google form link to answer the survey questionnaire. A total of fifty (50) people were included in the study's sample size. As shown in Table III, fifteen (15) from Central Philippines State University, ten (10) from Northern Negros State College of Science & Technology, fifteen (15) from Carlos Hilado Memorial State University, and ten (10) Technological University of the Philippines – Visayas.

A scale of 1 to 5 was used based on the Likert scale, and a non-comparative scaling approach was used.

The study expected a target of 50 respondents to answer the survey questionnaire. However, 35 of the estimated number filled out the surveys and returned them. Thus, as shown in Table III, the answer rate was 70%. This response rate was deemed suitable for analysis since, according to [14], a minimum of 70% or higher is exceptional for analysis.

TABLE III. DISTRIBUTION OF RESPONDENTS

SUC	Expected Responses	Percentage	Responses Received	Percentage
Central Philippines State University	15	30.00	15	42.90
Northern Negros State College of Science & Technology	10	20.00	5	14.29
Carlos Hilado Memorial State University	15	30.00	10	28.58
Technological University of the Philippines - Visayas	10	20.00	5	14.29
Total	50	100.00	35	100.00

#### V. DISCUSSION

##### A. Use of ICT

The purpose of the study was to ascertain the extent of the use of computers using ICT at the State Universities and Colleges (SUCs) in Negros Occidental, Philippines, which necessitated the creation of a business continuity plan for ICT. According to Table IV below, 0% of respondents did not rely on ICT for all their key operations, whereas 100% used ICT for all their essential duties. The value of information and communication technology (ICT) in education is unquestionable on a global scale. It has the potential to be very effective to use ICT to increase educational opportunities. ICT has the potential to improve the relevance and standard of education while expanding access to it [15].

TABLE IV. EXTENT OF ICT USAGE

Responses	Frequency	Percentage
Yes	50	100.00
No	0	0.00
Total	50	100.00

##### B. Failure of ICT core systems

The study aimed to ascertain how frequently the ICT systems SUCs relied on broke down. According to Table V below, 82.6% of respondents said their systems failed only occasionally, while 17.4% said it happened sometimes. 33.1% of the respondents reported that system failure-causing disasters never or only occasionally happened, compared to 45.6% who experienced disasters sparingly and 21.3% who never experienced system outages due to disasters. According to [16] it is crucial to improve their capability for ICT policy, respond to shifts in the ICT ecosystem, develop robust cybersecurity regulations, and make sure the private sector operates under predictable conditions.

TABLE V. FAILURE OF ICT CORE SYSTEMS

Responses' Percentage Distribution					
System Failures	Never	Rarely	Sometimes	Always	Total
Your systems fail frequently.	0.0%	82.6%	17.4%	0.0%	100%
Disasters occur, causing system outages	21.3%	45.6%	33.1%	0.0%	100%

C. Access to the ICT Continuity Plan

The study aimed to ascertain whether the four SUCs in the Philippine province of Negros Occidental had an ICT business continuity plan in place to lessen or avert the results of system malfunctions brought on by emergencies. Table VI demonstrates that only 2 SUCs had an ICT business continuity plan compared to the other 2 SUCs. Maintaining organizational operations in the face of potential threats, risks, the causes of power outages, cyberattacks, or epidemiological attacks, or natural disasters. BCP is therefore crucial in SUCs to plan for any risk [17].

TABLE VI. AVAILABILITY OF ICT CONTINUITY PLAN

Responses	Frequency	Percentage
Yes	2	50.00
No	2	50.00
Total	5	100.00

D. Availability of ICTCP Policies

The study looked into any internal guidelines for managing, regulating, and controlling the ICT business continuity plan. This was done to assess the level of implementation of ICT business continuity plans (Table VII). Two SUCs had established policies, while the other two did not.

TABLE VII. AVAILABILITY OF ICT BUSINESS CONTINUITY PLAN POLICIES

Responses	Frequency	Percentage
Yes	2	50.00
No	2	50.00
Total	4	100.00

E. Availability of Disaster Recovery Site

The study examined whether it was possible to check the effectiveness of the ICT business continuity plan at an off-site location for disaster recovery. One SUC lacked a recovery site, while the other three had them to help with recovery (Table VIII).

TABLE VIII. SITE FOR DISASTER RECOVERY IS ACCESSIBLE

Responses	Frequency	Percentage
Yes	3	75.00
No	1	25.00
Total	4	100.00

F. Challenges of ICTCP

The study aimed to identify the problems and develop an ICT business continuity strategy for the SUCs in the Philippine province of Negros Occidental. Table IX below shows that responses with a mean of 4.52 demonstrated strategies fail due to insufficient resources. With a standard of 3.80, respondents responded that the ICT plan requires an extensive planning for the second significant difficulty, and stakeholders in the firm do not have access to enough information concerning ICT business continuity plans. Limited resources were rated as the main obstacle to an ICT business continuity plan by 2.6. The institution's ICT strategy is not seen as a strategic component, according to respondents with a mean score of 2.41. According to respondents, with a mean of 2.57, the ICT business continuity plan is seen as an ICT-only job that excludes other departments. A standard of 2.5 respondents revealed that managing the ICT business continuity plan required technical competence and that stakeholders were not properly informed about what an ICT business continuity plan involved inside the organization.

TABLE IX. ICT BUSINESS CONTINUITY PLAN CHALLENGES

Items	N	Mean	Std. Deviation
ICT plans fail as a result of insufficient resources.	35	4.52	.936640
Plans for ICT business continuity need to be carefully thought out.	35	3.80	1.222376
The company's stakeholders are not adequately informed about what an ICT business continuity plan entails.	35	3.80	1.222376
The amount of time required to implement an ICT business continuity plan is excessive.	35	2.57	.654970
ICT business continuity plans cannot be updated as quickly as technology does.	35	2.41	.932764
ICT systems experience system failures more frequently than earlier manual processes, necessitating ongoing monitoring.	35	2.60	1.139540
The institution does not view the ICT business continuity plan as a strategic component.	35	2.41	.932764
The management of an ICT business continuity plan requires technical expertise.	35	2.57	.781736
There are no regulations in place to support the administration of the plan.	35	2.90	.987849
According to some, the ICT business continuity plan only pertains to ICT and excludes other departments.	35	2.57	.654970

G. Service Delivery

The study's goal was to ascertain how an ICT business continuity plan affected the provision of services. Table X demonstrates that the companies' profit margins increased as a result of customer satisfaction for the SUCs was based on the quality of services offered, offered superior goods and services with a mean of 3.57, and with a mean of 2.57, service delivery was boosted through the institution's adoption of an ICT business continuity plan.

TABLE X. SERVICE DELIVERY

Items	N	Mean	Std. Deviation
Customer satisfaction depends on how well services are provided.	35	3.57	1.184644
Service delivery is improved thanks to our company's adoption of an ICT business continuity plan.	35	2.57	.654970

## VI. CONCLUSION

The study found out the advantages of an ICT business continuity plan that most SUCs believed it benefited from the shortened time needed for system recovery after a failure. Other advantages were the development of teamwork and system understanding through employee engagement plans, the realization, and reduction of points of failure during the testing of the ICT business continuity plan, the ability to know what to do in the event of a disaster, reduction of losses caused by unforeseen disasters, effective resource planning for the company in the event of a disaster.

According to the study's findings, ICT has been widely adopted in SUCs, and it is clear that these institutions need their systems to operate continually with little chance of failure. Most insurance businesses implement an ICT business continuity plan to ensure fewer system disruptions and downtimes. Most SUCs have consistent ICT system uptime because they can continuously provide services; surprisingly, this is independent of an ICT business continuity plan. Most SUCs do not entirely implement their ICT business continuity plans. The ICT business continuity plan was easily overlooked or misunderstood as an ICT role because it had not been considered a strategic business need. It was also found to be resource-intensive and time- and planning-intensive.

The study's conclusions allow for the following recommendations: SUCs can significantly enhance service delivery if ICT business continuity planning is taken seriously, adopted, and entirely carried out. This is due to several reasons, including that they impact how services are delivered. An ICT business continuity strategy will ensure that ICT systems are always accessible to prevent service disruptions.

The proper management of ICT business continuity plans requires clearly defined rules, guidelines, and policies, efficient testing intervals, the availability of disaster recovery sites, enough time, hardware, software, and technology resources, as well as technical expertise and teamwork.

## ACKNOWLEDGMENT

The researchers would like to thank Dr. Felicisimo V. Wenceslao, Jr., professor of College of Information and Computing Studies at the Northern Iloilo State University, for introducing us to this topic and providing invaluable guidance and support during the research.

## REFERENCES

[1] S. Li and Y. Yan, "Data-driven shock impact of COVID-19 on the market financial system," *Inf Process Manag*, vol. 59, no. 1, Jan. 2022, doi: 10.1016/j.ipm.2021.102768.

[2] H. Kavonga, "ICT Business Continuity Plan And Service Delivery In Insurance Companies In Kenya A Project Submitted In Partial Fulfillment Of The Requirements For The Award Of The Degree Of Master Of Business Administration (Mba), School Of Business, University Of Nairobi 2017."

[3] M. Niemimaa, J. Järveläinen, M. Heikkilä, and J. Heikkilä, "Business continuity of business models: Evaluating the resilience of business models for contingencies," *Int J Inf Manage*, vol. 49, pp. 208–216, Dec. 2019, doi: 10.1016/j.ijinfomgt.2019.04.010.

[4] J. Hagelbäck, "Hybrid Pathfinding in StarCraft," *IEEE Trans Comput Intell AI Games*, vol. 8, no. 4, pp. 319–324, Dec. 2016, doi: 10.1109/TCIAIG.2015.2414447.

[5] M.-A. Kaufhold et al., "Business Continuity Management in Micro Enterprises: Perception, Strategies, and Use of ICT," *International Journal of Information Systems for Crisis Response and Management*, vol. 10, no. 1, pp. 1–19, Jan. 2018, doi: 10.4018/ijiscram.2018010101.

[6] E. D. Canedo et al., "Information and communication technology (ICT) governance processes: A case study," *Information (Switzerland)*, vol. 11, no. 10, pp. 1–28, Oct. 2020, doi: 10.3390/info11100462.

[7] S. A. Torabi, H. Rezaei Soufi, and N. Sahebjamnia, "A new framework for business impact analysis in business continuity management (with a case study)," *Saf Sci*, vol. 68, pp. 309–323, 2014, doi: 10.1016/j.ssci.2014.04.017.

[8] S. V. Fani and A. P. Subriadi, "Business continuity plan: Examining of multi-usable framework," in *Procedia Computer Science*, 2019, vol. 161, pp. 275–282. doi: 10.1016/j.procs.2019.11.124.

[9] E. Fasolis1, V. Vassalos2, and A. I. Kokkinaki3, "IFIP AICT 399 - Designing and Developing a Business Continuity Plan Based on Collective Intelligence," 2013.

[10] S. Fani and A. Subiadi, "Trend of Business Continuity Plan: A Systematic Literature Review," Feb. 2020. doi: 10.4108/eai.13-2-2019.2286164.

[11] N. Russo, L. Reis, C. Silveira, and H. S. Mamede, "Framework for designing Business Continuity-Multidisciplinary Evaluation of Organizational Maturity," in *Iberian Conference on Information Systems and Technologies*, CISTI, Jun. 2021. doi: 10.23919/CISTI52073.2021.9476297.

[12] G. Pramudya and A. N. Fajar, "Business Continuity Plan using ISO 22301:2012 in it Solution Company (PT. ABC)," in *It Solution Company (Pt. Abc) International Journal of Mechanical Engineering and Technology*, vol. 10, no. 2, pp. 865–872, 2019, [Online]. Available: <http://www.iaeme.com/IJMET/index.asp865http://www.iaeme.com/ijmet/issues.asp?JType=IJMET&VType=10&IType=2http://www.iaeme.com/ijmet/issues.asp?JType=IJMET&VType=10&IType=2>

[13] H. Rezaei Soufi, S. A. Torabi, and N. Sahebjamnia, "Developing a novel quantitative framework for business continuity planning," *Int J Prod Res*, vol. 57, no. 3, pp. 779–800, Feb. 2019, doi: 10.1080/00207543.2018.1483586.

[14] A. G. and A. M. Mugenda, "Qualitative research methods," 2013.

[15] K. Das, "International Journal of Innovative Studies in Sociology and Humanities (IJSSH) The Role and Impact of ICT in Improving the Quality of Education: An Overview," 2019, [Online]. Available: [www.ijssh.org](http://www.ijssh.org)

[16] T. Corrigan, "African perspectives Global insights Policy Briefing 197 Africa's ICT infrastructure: Its present and prospects," 2020. [Online]. Available: <https://www.afdb.org/en/knowledge/publications/tracking-africa%E2%8099>

[17] N. Roxana Moşteanu Professor, "Article ID: IJM\_11\_04\_018 Cite this Article: Dr. Narcisa Roxana Moşteanu, Management of Disaster and Business Continuity in a Digital World," *International Journal of Management*, vol. 11, no. 04, pp. 169–177, 2020, [Online]. Available: <http://www.iaeme.com/IJM/index.asp169http://www.iaeme.com/ijm/issue.asp?JType=IJM&VType=11&IType=4JournalImpactFactor>



# Towards a Machine Learning-based Model for Automated Crop Type Mapping

Asmae DAKIR, Fatimazahra BARRAMOU, Omar Bachir ALAMI

Team (SGEO) - Laboratory of Systems Engineering

Hassania School of Public Works (EHTP)

Casablanca, Morocco

**Abstract**— In the field of smart farming, automated crop type mapping is a challenging task to guarantee fast and automatic management of the agricultural sector. With the emergence of advanced technologies such as artificial intelligence and geospatial technologies, new concepts were developed to provide realistic solutions to precision agriculture. The present study aims to present a machine learning-based model for automated crop-type mapping with high accuracy. The proposed model is based on the use of both optical and radar satellite images for the classification of crop types with machine learning-based algorithms. Random Forest and Support Vector Machine, were employed to classify the time series of vegetation indices. Several indices extracted from both optical and radar data were calculated. Harmonical modelization was also applied to optical indices, and decomposed into harmonic terms to calculate the fitted values of the time series. The proposed model was implemented using the geospatial processing services of Google Earth Engine and tested with a case study with about 147 satellite images. The results show the annual variability of crops and allowed performing classifications and crop type mapping with accuracy that exceeds the performances of the other existing models.

**Keywords**—Smart farming; artificial intelligence; machine learning; precision agriculture; random forest; SVM

## I. INTRODUCTION

Agriculture has been a challenging economical sector, and a vital pillar of development in many countries. The first challenge is to ensure food self-sufficiency and respond to the increasing requirements of a growing population. The Food and Agriculture Organization of the United Nations [1] global report highlighted food insecurity and its accelerating rising trend. The report suggested that to prevent severe hunger, daring transformation must be conducted in the agri-food systems. Many recent technics and systems aimed to ensure the sustainable development of agriculture [2] and new concepts were developed like precision agriculture [3]. The recent revolution of digital technologies has radically changed agricultural management, like the employment of geospatial technologies, remote sensing data [4], and artificial intelligence [5], [6]. This study proposes a new model to show the contribution of machine learning algorithms in the

identification of crop types using both optical and radar satellite images with time series of different vegetation indices. The general aim is to establish a method for improving crop type mapping accuracies, with the demonstration of the contribution of optical and radar data and their complementarity.

The paper proceeds as follows: providing related works in Section II. Section III, introduces the proposed model. The case study is presented in Section IV. Section V presents and discusses the results, and Section VI concludes and gives the intended future works.

## II. RELATED WORKS

Crop-type mapping using remote sensing data and machine learning technics is the subject of multiple research [7]–[10]. Optical data has been a reference data for crop mapping studies because of the ease to link phenological development and biological properties of crops with optical acquisitions to differentiate crop species [11]–[13]. Also, it identifies various growing stages of a single crop, rice as in this case study [14]. The author in [11] used Sentinel-2 time series data with gap filling method to overcome data discontinuity caused by cloud cover. Interpolation technics were also used by [11] in the DATimes software to capture seasonal vegetation dynamics. Different optical sensors were also combined to increase time series temporal frequency and to catch field-level phenologies [15]. Both optical and radar data were used by [16] to detect paddy rice fields using phenological variations and a textural-based strategy. Radar images only were investigated by [17] to detect winter wheat phenological stages. They analyzed the temporal variations of the Sentinel -1 time series in the function of different phenological phases. The author in [18] as well used Sentinel-1 time series to conduct classification considering Spatiotemporal phenological information.

Different machine-learning algorithms were employed to produce accurate crop maps. The support vector machine (SVM), and random forest (RF) classifiers have been the most popular in recent years for the classification of satellite images [19]. Many papers reported better performance of SVM [8], [20]–[22] as well as DT and RF algorithms compared to other techniques.

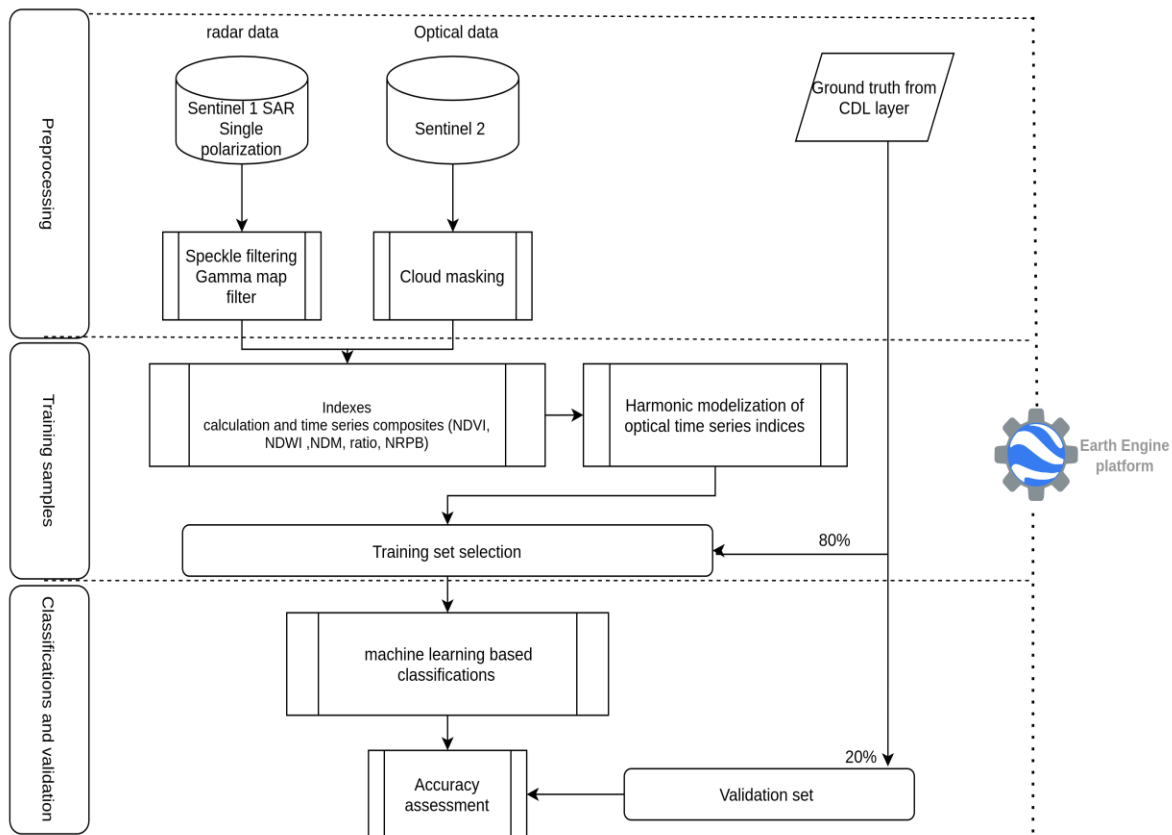


Fig. 1. Proposed model

The goal of mapping crops was approached using different methods, technics and datasets in the literature review. Although the obtained results presented good accuracies, the development of geospatial technics has imported its evolution; the integration of the Google Earth Engine allows producing large treatments instantly that can be employed to identify crops early in the season. The obtained accuracies in the state of the art are yet to be improved by exploiting complementarity between radar and optical time-series and their indices.

This research proposes a new machine-learning method combining time-series indices extracted from optical and radar satellite images and machine-learning classifiers to perform automated and high-accuracy crop-type mapping.

### III. PROPOSED MODEL

The proposed model employs time series from both optical and radar data. The model was performed using the geospatial processing services of Google Earth Engine (Fig. 1):

#### A. Preprocessing

**Cloud masking:** An important step of pre-processing the image collection is to omit the disturbance caused by clouds and shadows from the imagery. The cloud masking process was performed using the cloud probability band that was created with the Sentinel 2 cloud detector library. The maximum cloud probability was limited to 25. The gaps in the masked image were then filled with the previous interpolated image.

**Speckle filtering:** satellite images are usually affected by speckle noise. Multiple statistical methods were developed to remove the speckle in the concern to preserve image details. The study conducted by [23] compared different filtering methods dedicated to speckle suppression in SAR images and found that Lee-Sigma and Gamma-MAP are showing relatively good detail preserving abilities than other filter types. The author in [24] also concluded that the Gamma Map filter is reliable as proved by the comparison between the Lee filter, frost filter, and Gamma Map. In the present model, the Gamma-MAP filter is used, which is based on the Bayesian analysis of image statistics. It uses the Maximum A-Posteriori (MAP) estimation method. While using this filter, Gamma distribution is assumed for the underlying image and the speckle noise in it. Thus, this filter works best for geospatial images containing homogenous areas such as oceans, forests, fields, etc.

#### B. Training Sample

1) **Indices calculation and time series composites:** Optical data provide information in multiple bands that can produce valuable information about the state of vegetation. For the purpose to capture spatiotemporal variation in photosynthetically active vegetation, multiple optical indices were developed in the literature to characterize and monitor the development of crops [19], [25]. The author in [25] calculated the EVI and NDVI indices from the time series to extract metrics for crop discrimination. In this study, different

indices were calculated for each image in the image collection (Table I). The main used bands are Red (R), Green (G), Near InfraRed (NIR), and Short-Wavelength InfraRed (SWIR) from the optical images, and both polarization VV and VH from radar images.

TABLE I. OPTICAL AND BACKSCATTER CALCULATED INDICES

Index	Abbreviation	Formula
Normalized Difference Vegetation Index	NDVI	$NDVI = \frac{NIR-R}{NIR+R} \quad (1)$
Normalized Difference Water Index	NDWI	$NDWI = \frac{G-NIR}{G+NIR} \quad (2)$
Normalized Difference Moisture Index	NDMI	$NDMI = \frac{NIR-SWIR}{NIR+SWIR} \quad (3)$
Normalized Ratio Procedure between Bands	NRPB	$NRPB = \frac{\sigma_{VH} - \sigma_{VV}}{\sigma_{VH} + \sigma_{VV}} \quad (4)$
The polarization ratio	Ratio	$Ratio = \frac{\sigma_{VV}}{\sigma_{VH}} \quad (5)$

2) *Harmonic modelization of optical time series*: Time series from the optical indices depends on the phenological cycle of crops throughout the year. The analysis of the variations is represented by applying harmonic modeling also named Fourier analysis. The analysis consists of decomposing the time-dependent periodic event into a series of sinusoidal functions, with phase and amplitude values. The general equation of a time series is presented by [26] in eq. (6).

$$A \cos(2\pi\omega t - \varphi) = \beta_2 \cos(2\pi\omega t) + \beta_3 \sin(2\pi\omega t) \quad (6)$$

Considering the linear model, where A is amplitude,  $\omega$  is a random error,  $\omega$  is frequency, and  $\varphi$  is phase:

$$P_t = NDVI_t = \beta_0 + \beta_1 t + A \cos(2\pi\omega t - \varphi) + \epsilon_t$$

$$= \beta_0 + \beta_1 t + \beta_2 \cos(2\pi\omega t) + \beta_3 \sin(2\pi\omega t) + \epsilon_t$$

$$\text{And } \beta_2 = A \cos(\varphi) \quad \beta_3 = A \sin(\varphi)$$

$$A = \text{amplitude} = (\beta_2^2 + \beta_3^2)^{1/2} \quad \varphi = \text{phase} = \text{atan}(\beta_3/\beta_2)$$

For each optical index, the harmonic modeling was then applied, and time series were decomposed into harmonic terms to calculate the fitted values of the time series.

3) *Training set selection*: To guarantee a good presentation of each class, training samples should respect a good representation of each class taking into consideration spatial distribution. 20% of the samples are set for validation and accuracy calculations, and 80% were used for training and extraction features from the formulated time series.

### C. Classification and Validation

In the literature, different classification methods are employed for land cover and land mapping. This study, employed two classifiers which are the most performant [8], [20]–[22].

1) *Machine learning classification*: Random Forest RF classifier is based on building multiple trees from samples of the training data. Each tree is built using a different subset from the original training variables. Its advantage is that the algorithm can handle a huge amount of input data. The decision of belonging to a given class is determined by the majority vote of the trees.

Support Vector Machine (SVM) is a supervised non-parametric statistical technique. The decision to separate between classes is made by calculating the hyperplane that maximizes the margin between classes. The separation between data points is based on the applied kernel function (Linear, Polynomial, Gaussian, Radial Basis Function (RBF), or Sigmoid) that determines the efficiency of the classification.

2) *Accuracy assessment*: Two performance criteria were used to assess the result's accuracy. The main index of Cohen's kappa is a statistical measure of interrater reliability for categorical variables. It takes into account the possibility of the accord occurring by chance (eq. 7).

$$\frac{p_0 - pc}{1 - pc} \quad (7)$$

While  $p_0$  = Observed accuracy.  $p$  is the sum of relative frequency in the diagonal of the error matrix.  $pc$  = Chance agreement.

F1 score also a measure of a model's accuracy can be interpreted as a harmonic mean of the precision and recall of the confusion matrix. F1-score is calculated per class for a multiclass classification problem (eq. 8).

$$F1 \text{ Score} = \frac{2}{\frac{1}{\frac{Recall}{precision}} + 1} = \frac{2 * Precision * Recall}{Precision + Recall} \quad (8)$$

$$\text{Where: recall} = \frac{Tp}{Tp + Fn} \text{ and precision} = \frac{Tp}{Tp + Fp}$$

## IV. CASE STUDY

### A. Study Area

To evaluate the radar and optical indices using a supervised classification method, the proposed model is tested in an agricultural zone in Minnesota State in the United States (Fig. 2). Minnesota is located in the Western part of the Great Lakes region and ranks fifth in the United States for total crop sales, the major crops are corn, Soybean, sugar beets, and dry beans.

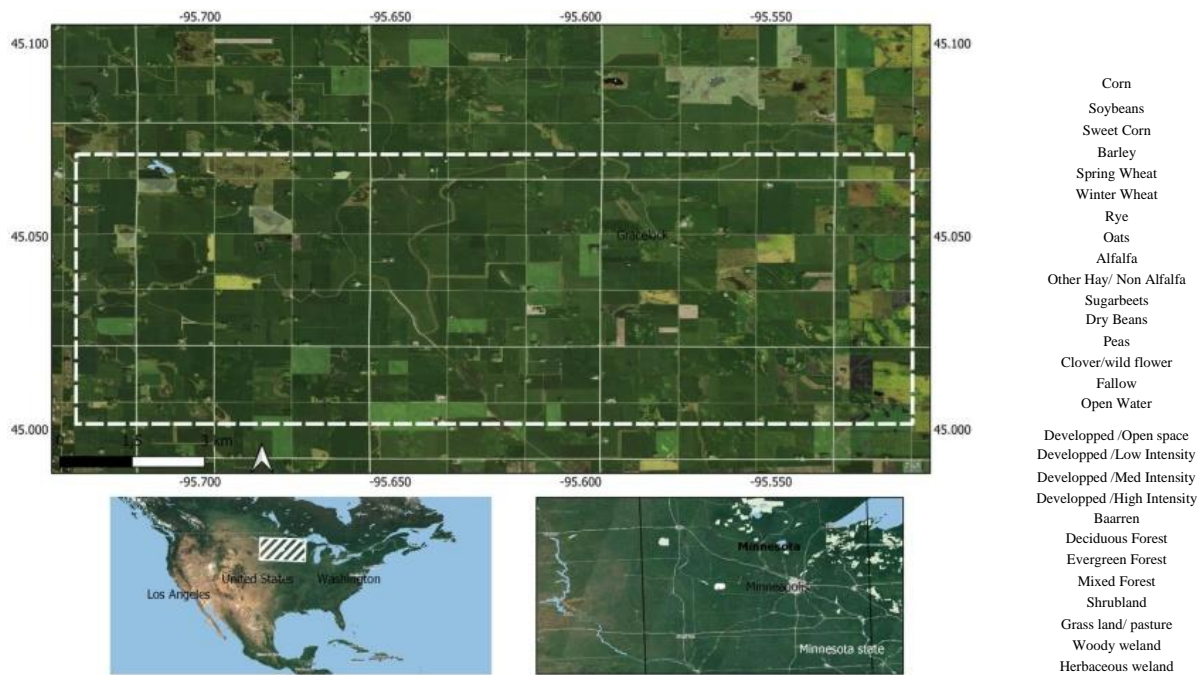


Fig. 2. Study area in the Minnesota state and the studied classes

### B. Crop Inventory Data

The reference data were collected from the cropland data layer (CDL) produced by United States Department of Agriculture. The layer contains annual crops from extensive agricultural ground truth with 30m spatial resolution. The process of training started with random points selection taking into consideration to cover the totality of the study area, and covering all the agricultural types. 80 % of data presenting 5989 points were selected for training the model, and 20% for validation. The selected zone contains 14 types of crops with 4 major types.

### C. Optical and Radar Data

Both radar (Sentinel 1) and optical (Sentinel 2) images were used in this study. The Sentinel 1 mission provides C-band Synthetic Aperture Radar data (SAR). The image catalog of Sentinel 1 data provides preprocessed images, terrain corrected and radiometrically calibrated. A total of 29 scenes of Synthetic Aperture Radar images were used from 01-01-2019 to 30-12-2019. The images were restricted to single-polarization VV and VH. The active sensor expands the possibilities of acquiring data in cloudy weather allowing then better monitoring of the vegetation evolution.

The Sentinel-2 mission provides multispectral high-resolution imagery with 12 spectral bands. The image collection contains 145 optical images covering all the studied periods.

### D. Analysis Platform

The development of the remote sensing field imported different offers and a large amount of data from different sensors, and several platforms have been elaborated to handle geospatial analysis and processing. The Google Earth Engine was introduced as a multi-petabyte catalog and cloud computing platform with high-performance computation

capabilities and has been investigated in land cover studies [27] and agricultural studies [28], [29].

The proposed process was all performed in the Google Earth Engine (GEE) platform, from the Sentinel image selection to the validation process. The GEE platform allowed the process of large-density images for pixel-based image analysis as well as the classification algorithms due to the high cloud calculation performance the platform offers.

### E. Time Series Formulation, Training, and Machine Learning Algorithms

The first steps of processing time series are conducted as detailed in the previous section. After the preprocessing, calculating optical and radar indices of each imagery data was performed.

SAR indices were extracted from the single-polarization bands. The Normalized Ratio Procedure between Bands (NRPB) and the ratio were estimated using the equation in Table I where  $\sigma_{VH}$  and  $\sigma_{VV}$  are the backscatter VH and VV polarization. In the same way, the NDVI, NDMI, and NDWI optical indices were calculated, then applied the harmonic modelization of the time series.

The training was then applied to the formulated input, the training set was selected randomly from the time series stack generated from all the SAR and optical calculated indices. The input features are then fed to the employed machine learning-based classifiers.

Multiple parameters were tested for obtaining perfect results. The final parameters for the SVM classifier were set to the Radial Basis Function (RBF), 0.5 for gamma and 10 for the cost. Random Forest is applied using 800 trees and 20 variables per split.

## V. RESULTS AND DISCUSSION

### A. Time Series of Vegetation Indices

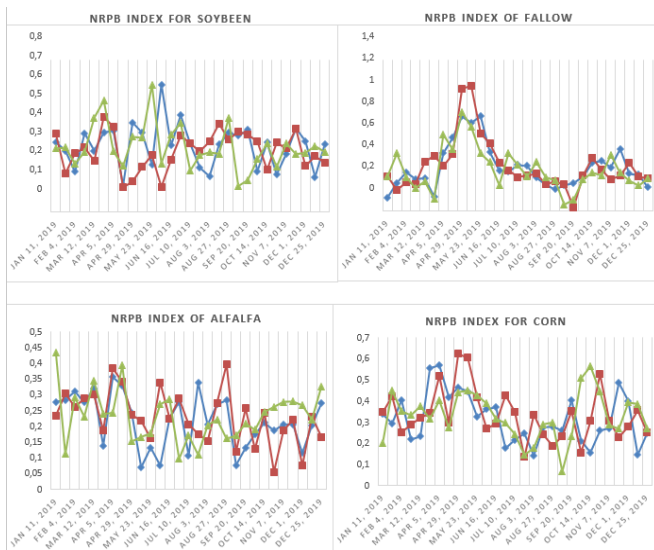


Fig. 3. NRPB index time series of major crops in the studied zone

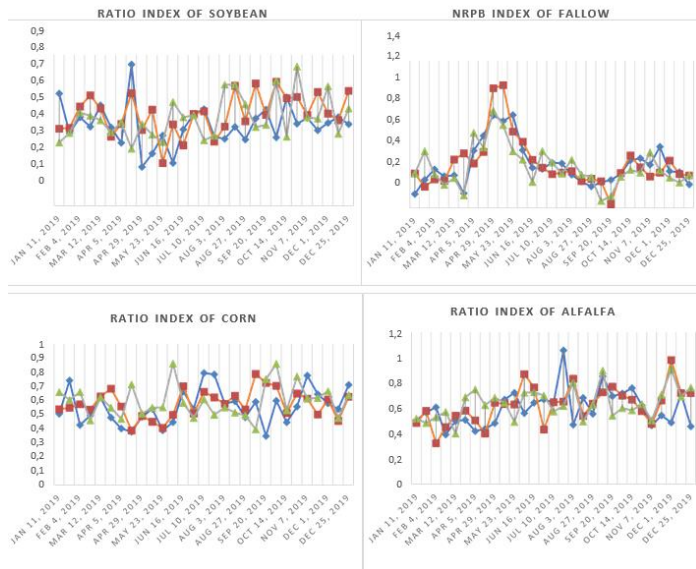


Fig. 4. Ratio index time series of major crops in the studied zone

From the SAR indices, the NRPB and ratio index time series were presented in Fig. 3 and Fig. 4 with a selection of 4 major crops. The time series of the Normalized Ratio Procedure between Bands and the ratio index can monitor the vegetation changes. The temporal signature of different considered crop types is showing different signature behavior. The NRPB time series of corn alfalfa and soybean know significant variations whereas Fallow parcels responded smoothly and monotonously to changes over the year.

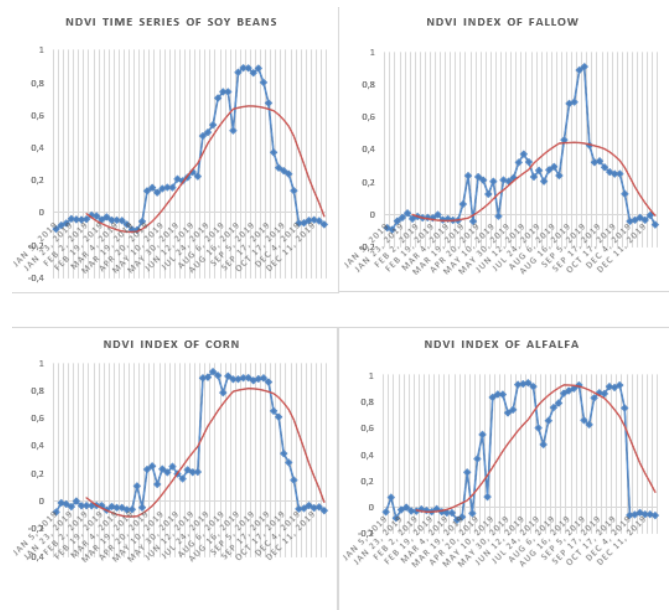


Fig. 5. NDVI time series and fitted values of major crops in the studied zone

The variations are a function of soil surface conditions, moisture, roughness, and biomass development of crops. The NRPB index was also used by [30] to generate metrics for the input set of the model to aid the prediction of NDVI and highlighted the similarity found in the NDVI and  $\sigma_{VH}/\sigma_{VV}$  ratios with crops and finally found that insertion of the NRPB variable in machine learning models, like RF, gives better results.

Times series of the NDVI allows the characterization of each crop. Since the NDVI is a perfect index to describe the chlorophyll activity of crops, a dense and healthy state is presented by a high value of the NDVI index reaching 1, in the opposite case, the value approaches 0. Then, the time series is presenting the phenological cycle of each crop.

Fig. 5 presents a selection of 4 crops of the NDVI time series in the studied area and the harmonic modeling values. The model is suitable for smoothing the spectral curves and allows distinction between crops. Corn, dry beans, and soybeans are presenting a unimodal periodic model, with a high value for corn culture. The resulting phenological cycles match the phenological calendar provided by the USDA National Agricultural Statistics. Corn starts in late April and is harvested in early November. While the phenological cycle of Soybean Starts with the plantation in early May and is harvested in late October. Alfalfa is presenting the highest amplitude values and a different curve from other crops due to agricultural practices. Alfalfa is harvested repeatedly during the growing season starting from early April to late October. The results of the obtained phenological cycles were compatible with the crop calendar as given by the USDA in the region of Minnesota.

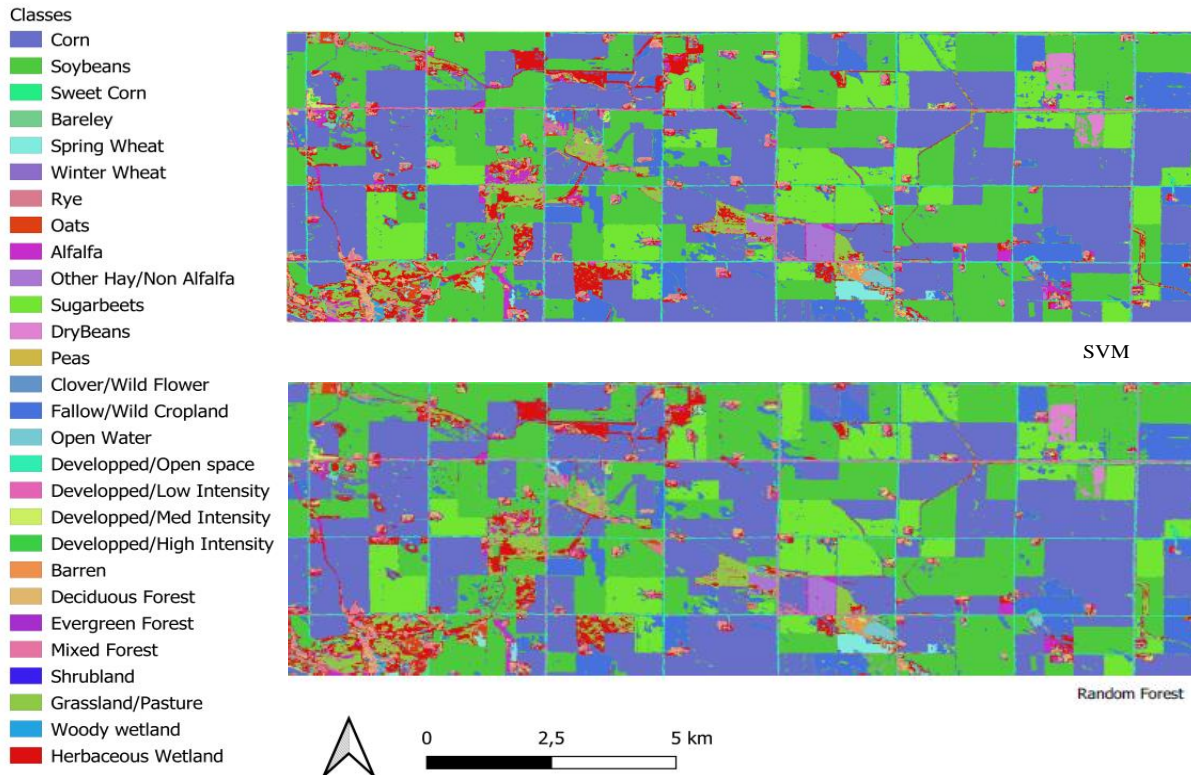


Fig. 6. Crop type mapping: Final results of RF and SVM classifications.

### B. Classification Results

SVM and Random Forest have demonstrated their advantage in classifying agricultural cover maps. The validation of the classification results was conducted by calculating the confusion matrix. 20% of random samples were used to validate the final result. Both classifiers had given good results with the advantage of the random forest classifier with 0.95 kappa index, and 0.85 for SVM. Table II presents accuracy metrics with good accuracy results, with the advantage of the RF classifier. Other studies had demonstrated the complementarity of optical and radar data [20], [31]–[34]. The authors in [35] have found an overall accuracy of 93.83% from combined inputs. The authors in [8] have found an overall accuracy between 73% and 95% depending on the input dataset used, using the SVM classifier.

Fig. 6 presents the final classification using SVM and RF algorithms. The results are showing similar classes except for slight visual differences.

Performances metrics were calculated, other than the kappa index, the F1-score, and Producer accuracy are presented in Table II. The Producer accuracy represents the probability that a particular sample of a particular class is classified correctly. The most correctly attributed classes are barley, corn, soybeans, and winter wheat.

### VI. CONCLUSION

This research study deals with the problem of crop type identification. A machine learning-based model for automated crop type mapping is proposed. The novelty of the model is to improve crop type mapping accuracy using time series from both optical and radar images by extracting vegetation indices. The model presents several advantages. It demonstrate the complementarity between optical and radar satellite images for crop type mapping studies. Secondly, the results pointed the advantage of Random Forest classifier over SVM. The resulted accuracy outperformed existing models in the state of the art with a kappa index of 95%.

The proposed model was implemented using Google Earth Engine and tested with a specific case study in an agricultural zone in Minnesota State in the United States. Future works intend to assess and compare the performances of deep learning and machine learning algorithms for crop-type mapping.

### ACKNOWLEDGMENT

We are very thankful to Unites State Department of Agriculture (USDA), and the National Agricultural Statistics Service (NASS) for providing the Crop Data Layer from which crop types were identified.

TABLE II. VALIDATION ACCURACY METRICS OF RANDOM FOREST AND SVM CLASSIFICATIONS

Classes	F1-SCORE RF	F1-SCORE SVM	Producer accuracy RF	Producer accuracy SVM
Corn	0.97	0.72	0.98	1
Soybeans	0.96	0.89	0.97	0.78
Sweet Corn	0.95	0.85	0.92	0.76
Barley	1	0.89	1	0.69
Spring Wheat	0.95	0.87	0.93	0.84
Winter Wheat	0.96	0.85	0.96	0.7
Rye	0.96	0.87	0.92	0.75
Oats	0.96	0.87	0.97	0.69
Alfalfa	0.88	0.95	1	0.7
Other Hay/ Non Alfalfa	0.87	0.87	0.87	0.79
Sugarbeets	0.96	0.91	0.94	0.8
Dry Beans	0.96	0.89	0.97	0.84
Peas	0.95	0.93	0.88	0.87
Clover/wildflower	0.85	0.66	1	0.7
Fallow	0.95	0.93	0.96	0.77
Open Water	0.99	0.89	0.97	0.69
Developed /Open space	0.94	0.90	0.94	0.86
Developed /Low Intensity	0.95	0.85	0.95	0.77
Developed /Med Intensity	0.89	0.98	0.94	0.94
Developed /High Intensity	0.83	0.85	0.8	0.82
Baarren	0.88	0.79	0.84	0.75
Deciduous Forest	0.97	0.85	1	0.78
Evergreen Forest	0.90	0.88	0.86	0.80
Mixed Forest	0.89	0.91	0.88	0.81
Shrubland	0.91	0.88	0.93	0.82
Grassland/pasture	1	0.9	0.86	0.75
Woody wetland	0.90	0.93	0.94	0.83
Herbaceous wetland	0.90	0.80	0.93	0.75

REFERENCES

[1] FAO, The State of Food and Agriculture 2021: Making agrifood systems more resilient to shocks and stresses. Rome, Italy.; FAO, 2021. Doi: 10.4060/cb4476en.

[2] M. Shepherd, J. A. Turner, B. Small, et D. Wheeler, « Priorities for science to overcome hurdles thwarting the full promise of the ‘digital agriculture’ revolution », *J. Sci. Food Agric.*, vol. 100, no 14, p. 5083-5092, 2020, doi: <https://doi.org/10.1002/jsfa.9346>.

[3] S. Bagwari, « Impact of Internet of Things Based Monitoring and Prediction System In precision Agriculture », vol. 22, p. 4599, août 2019.

[4] S. Wang, G. Azzari, et D. B. Lobell, « Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering

techniques », *Remote Sens. Environ.*, vol. 222, p. 303-317, mars 2019, doi: 10.1016/j.rse.2018.12.026.

[5] A. Dakir, F. Barramou, et O. Alami, « Opportunities for Artificial Intelligence in Precision Agriculture Using Satellite Remote Sensing », 2022, p. 107-117. Doi: 10.1007/978-3-030-80458-9\_8.

[6] D. Lary, « Artificial Intelligence in Geoscience and Remote Sensing », 2010. Doi: 10.5772/9104.

[7] E. Amin et al., « The Sensagri Sentinel-2 Lai Green and Brown Product: from Algorithm Development Towards Operational Mapping », in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, juill. 2018, p. 1822-1825. Doi: 10.1109/IGARSS.2018.8518938.

[8] I. Aneece et P. Thenkabil, « Accuracies Achieved in Classifying Five Leading World Crop Types and their Growth Stages Using Optimal Earth Observing-1 Hyperion Hyperspectral Narrowbands on Google Earth Engine », *Remote Sens.*, vol. 10, no 12, p. 2027, déc. 2018, doi: 10.3390/rs10122027.

[9] A. Asgarian, A. Soffianian, et S. Pourmanafi, « Crop type mapping in a highly fragmented and heterogeneous agricultural landscape: A case of central Iran using multi-temporal Landsat 8 imagery », *Comput. Electron. Agric.*, vol. 127, p. 531-540, sept. 2016, doi: 10.1016/j.compag.2016.07.019.

[10] A. Bailly, « Time Series Classification Algorithms with Applications in Remote Sensing », p. 181, 2018.

[11] S. Belda et al., « datimes: A machine learning time series GUI toolbox for gap-filling and vegetation phenology trends detection », *Environ. Model. Softw.*, vol. 127, p. 104666, mai 2020, doi: 10.1016/j.envsoft.2020.104666.

[12] R. Luciani, G. Laneve, et C. A. Silva, « Crop Fields Classification Based on in Situ Phenological Metrics », in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, juill. 2019, p. 6306-6309. Doi: 10.1109/IGARSS.2019.8900444.

[13] H. Do Nascimento Bendini et al., « Comparing Phenometrics Extracted From Dense Landsat-Like Image Time Series for Crop Classification », in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, juill. 2019, p. 469-472. Doi: 10.1109/IGARSS.2019.8898139.

[14] S. Guha, T. Pal, et V. Ravibabu Mandla, « Crop Phenology Study Based on Multispectral Remote Sensing », in *Lecture Notes in Civil Engineering*, 2019, p. 985-993. Doi: 10.1007/978-981-10-8016-6\_68.

[15] P. Griffiths, C. Nendel, et P. Hostert, « Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping », *Remote Sens. Environ.*, vol. 220, p. 135-151, janv. 2019, doi: 10.1016/j.rse.2018.10.031.

[16] M. Singha et S. Sarmah, « Incorporating crop phenological trajectory and texture for paddy rice detection with time series MODIS, HJ-1A and ALOS PALSAR imagery », *Eur. J. Remote Sens.*, vol. 52, no 1, p. 73-87, janv. 2019, doi: 10.1080/22797254.2018.1556568.

[17] A. Nasrallah et al., « Sentinel-1 Data for Winter Wheat Phenology Monitoring and Mapping », *Remote Sens.*, vol. 11, no 19, p. 2228, sept. 2019, doi: 10.3390/rs11192228.

[18] B. Kenduywo, D. Bargiel, et U. Soergel, « Crop-type mapping from a sequence of Sentinel 1 images », *Int. J. Remote Sens.*, p. 1-22, avr. 2018, doi: 10.1080/01431161.2018.1460503.

[19] A. Dakir, B. Omar, et B. Fatimazahra, Crop type mapping using optical and radar images: a review. 2020, p. 8. Doi: 10.1109/Morgeo49228.2020.9121869.

[20] S. Park et J. Im, « Classification of croplands through fusion of optical and sar time series data », *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLI-B7, p. 703-704, juin 2016, doi: 10.5194/isprs-archives-XLI-B7-703-2016.

[21] U. Lussem, C. Hütt, et G. Waldhoff, « Combined analysis of sentinel-1 and rapideye data for improved crop type classification: an early season approach for rapeseed and cereals », *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLI-B8, p. 959-963, juin 2016, doi: 10.5194/isprs-archives-XLI-B8-959-2016.

[22] J. K. Gilbertson, J. Kemp, et A. Van Niekerk, « Effect of pan-sharpening multi-temporal Landsat 8 imagery for crop type differentiation using

- different classification techniques », *Comput. Electron. Agric.*, vol. 134, p. 151-159, mars 2017, doi: 10.1016/j.compag.2016.12.006.
- [23] P. Kupidura, « Comparison of filters dedicated to speckle suppression in sar images », *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLI-B7, p. 269-276, juin 2016, doi: 10.5194/isprsarchives-XLI-B7-269-2016.
- [24] J. Senthilnath, V. Handiru, R. Rajendra, O. S N, V. Mani, et P. Diwakar, « Integration of speckle de-noising and image segmentation using Synthetic Aperture Radar image for flood extent extraction », *J. Earth Syst. Sci.*, vol. 122, p. 559-572, juin 2013, doi: 10.1007/s12040-013-0305-z.
- [25] D. Arvor, M. Jonathan, M. Simoes, V. Dubreuil, et L. Durieux, « Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil », *Int. J. Remote Sens.*, vol. 32, p. 7847-7871, nov. 2011, doi: 10.1080/01431161.2010.531783.
- [26] R. H. Shumway et D. S. Stoffer, *Time Series Analysis and Its Applications: With R Examples*, 4e éd. Springer International Publishing, 2017. Doi: 10.1007/978-3-319-52452-8.
- [27] L. Kumar et O. Mutanga, *Google Earth Engine Applications*. MDPI, Basel, 2019. Doi: 10.3390/books978-3-03897-885-5.
- [28] T.-N. Phan, V. Kuch, et L. Lehnert, « Land Cover Classification using Google Earth Engine and Random Forest Classifier - The Role of Image Composition », *Remote Sens.*, juill. 2020, doi: 10.3390/rs12152411.
- [29] S. Xie, L. Liu, X. Zhang, J. Yang, X. Chen, et Y. Gao, « Automatic Land-Cover Mapping using Landsat Time-Series Data based on Google Earth Engine », *Remote Sens.*, vol. 11, p. 3023, déc. 2019, doi: 10.3390/rs11243023.
- [30] R. Filgueiras, E. C. Mantovani, D. Althoff, E. I. Fernandes Filho, et F. F. Da Cunha, « Crop NDVI Monitoring Based on Sentinel 1 », *Remote Sens.*, vol. 11, no 12, Art. No 12, janv. 2019, doi: 10.3390/rs11121441.
- [31] S. Park, J. Im, S. Park, C. Yoo, H. Han, et J. Rhee, « Classification and Mapping of Paddy Rice by Combining Landsat and SAR Time Series Data », *Remote Sens.*, vol. 10, p. 447, mars 2018, doi: 10.3390/rs10030447.
- [32] L. Mansaray, W. Huang, D. Zhang, J. Huang, et J. Li, « Mapping Rice Fields in Urban Shanghai, Southeast China, Using Sentinel-1A and Landsat 8 Datasets », *Remote Sens.*, vol. 9, p. 257, mars 2017, doi: 10.3390/rs9030257.
- [33] S. Giordano, S. Bailly, L. Landrieu, et N. Chehata, « Temporal Structured Classification of Sentinel 1 and 2 Time Series for Crop Type Mapping », juillet 2018. Consulté le: 16 août 2019. [En ligne]. Disponible sur: <https://hal.archives-ouvertes.fr/hal-01844619>
- [34] B. Salehi, B. Daneshfar, et A. M. Davidson, « Accurate crop-type classification using multi-temporal optical and multi-polarization SAR data in an object-based image analysis framework », *Int. J. Remote Sens.*, vol. 38, no 14, p. 4130-4155, juill. 2017, doi: 10.1080/01431161.2017.1317933.
- [35] J. D. T. De Alban, G. M. Connette, P. Oswald, et E. L. Webb, « Combined Landsat and L-Band SAR Data Improves Land Cover Classification and Change Detection in Dynamic Tropical Landscapes », *Remote Sens.*, vol. 10, no 2, p. 306, févr. 2018, doi: 10.3390/rs10020306.



# A Hybrid Model by Combining Discrete Cosine Transform and Deep Learning for Children Fingerprint Identification

Vaishali Kamble<sup>1</sup>, Manisha Dale<sup>2</sup>, Vinayak Bairagi<sup>3</sup>

Department of Electronics and Telecommunication, AISSMS Institute of Information Technology, Pune<sup>1</sup>

Department of Electronics and Telecommunication, Modern Education Society's College of Engineering, Pune, India<sup>1</sup>

Department of Electronics and Telecommunication, Modern Education Society's College of Engineering, Pune, India<sup>2</sup>

Department of Electronics and Telecommunication, AISSMS Institute of Information Technology, Pune<sup>3</sup>

**Abstract**—Fingerprint biometric as an identification tool for children recognition was started in the late 19<sup>th</sup> century by Sir Galton. However, it is still not matured for children as adult fingerprint identification even after the span of two centuries. There is an increasing need for biometric identification of children because more than one million children are missing every year as per the report of International Centre of missing and exploited children. This paper presents a robust method of children identification by combining Discrete Cosine Transform (DCT) features and machine learning classifiers with Deep learning algorithms. The handcrafted features of fingerprint are extracted using DCT coefficient's mid and high frequency bands. Gaussian Naïve Base (GNB) classifier is best fitted among machine learning classifiers to find the match score between training and testing images. Further, the Transfer learning model is used to extract the deep features and to get the identification score. To make the model robust and accurate score level fusion of both the models is performed. The proposed model is validated on two publicly available fingerprint databases of children named as CMBD and NITG databases and it is compared with state-of-the-art methods. The rank-1 identification accuracy obtained with the proposed method is 99 %, which is remarkable compared to the literature.

**Keywords**—Discrete Cosine Transform (DCT); Curve DCT; biometric recognition; machine learning; convolutional neural network; AlexNet

## I. INTRODUCTION

Children missing, swapping and abduction is caused due to non-recognition of them from their faces or other biometrics as they are in the developing stage in this age. The rate of mishaps happening to them is increasing day by day due to non-availability of the authentic system for recognition of them. Researchers studied the different biometrics like face [1,2,3,4], palmprint [5], footprint [6], ear [7] and headprints [8] for recognition of children.

Fingerprint modality is most widely used biometric for recognition of adults. Dr. Faulds [9] used fingerprints for identification of persons in 1870. As per Sir Galton theory, the fingerprint similarity chance in two different persons is 1 in 64 billion [10]. In children, the fingerprints show distinctive features which can differentiate them from others, though they are in the developing stage.

Fingerprint recognition can be done by two methods: minutiae matching and pattern matching. Minutiae are small details in fingerprints like ridge, valley, ridge ending, bifurcation. In minutiae matching, different minutiae are matched against each other and the highest match probability is considered as the correct match. The problem with minutia match for children recognition is biometric aging. The displacement of minutiae points due to growth in fingers. One needs a growth model to recognise the longitudinal (after time lapse) fingerprint images of children. The second method of pattern matching is based on the texture of fingerprint images. Fingerprint images are rich in texture and hence frequency domain representation is suited for it. The basic block diagram of fingerprint recognition is as shown in Fig. 1. Original fingerprint is processed and enhanced to get the most suitable features. The features are extracted using the feature extraction algorithms. These features are stored as feature vectors in the training process. The test phase is similar up to feature extraction, then these features are compared with stored features with the help of a matching algorithm or by using a classifier. The estimated match with the test image is the output of the identification system.

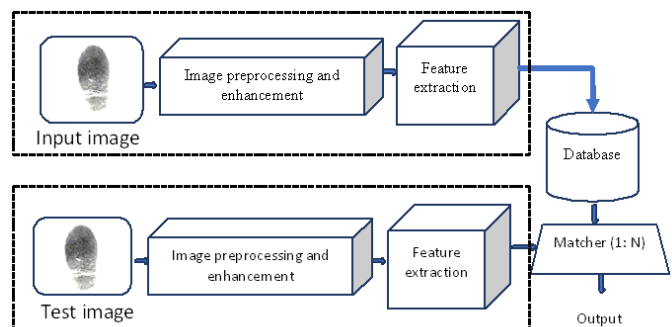


Fig. 1. Basic block diagram of image identification

## A. Literature Review

Automatic recognition of children using fingerprints was first studied by Jain et al. [11] in 2014 to increase the coverage of vaccination for the age group of 0 to 4 years. They used the up-sampling process by a factor of 1.8 to match the finger ridge distance to adults. Camacho et al. [12] also proposed a solution of up-sampling the fingerprint by an interpolation

factor based on children's ridge distance. Preciozzi et al. [13] formulated a scale factor based on distance between the ridges of adults to the distance between ridges of an age of child. By using these scale factors, the verification accuracy of children's fingerprints increases. Haraksim et al. [14] computed fingerprint minutiae-based growth model to reduce the biometric aging result on children recognition

Further, the hardware implementation of the fingerprint systems is also done to verify the accuracy of the fingerprints recognition. Koda et al. [15], designed the high resolution (1270 ppi) fingerprint scanner for scanning the minute details of children's fingerprints. Kalisky et al. [16] also designed high resolution platen free optical scanner. Engelsma et al. [17] studied the infants of 0 to 3 months to reduce the mortality of infants. They used a 1900 ppi scanner to capture the details of the infant's fingerprint. Improvement in the accuracy is observed by increasing resolution of the scanner. Macharia et al. [18] showed the Android based system for recognition of children and discussed the fingerprint quality of children. Engelsma et al. [19] further studied the problem of recognition of children for vaccination and nutritional supply with the 1900 ppi RaspiReader designed by him.

Most of the researchers used Commercial Software Defined Kits (SDK) for the recognition of children [20,21]. Jain et al. [22] continued his research in children recognition and this time collected the longitudinal database of infants. CNN is used for improving the quality of images. Feature extraction and matching is done with Commercial SDK. The longitudinal study is continued by Jain et al [23] for children verification. They used an Automatic fingerprint recognition system for feature extraction and matching. Engelsma et al. [19] extracted features using a texture based Convolutional Neural Network (CNN) matcher with two Commercial off the shelf matchers (COTS). Patil et al. [24] designed a fingerprint recognition system for infants and toddlers by extracting the finger codes by Gabor filtering and matching them using Euclidean distance. Moolla et al [25] investigated for the best modality for infants' recognition among fingerprint, iris and ear. The Research done in children fingerprint recognition is very less. The publicly available databases of children's fingerprints are also less. It is observed that feature extraction and matching of children's fingerprints is majorly done using commercial recognition systems. In latest research CNN and COTS systems are studied and are combined to achieve high accuracy. Transform based features are rarely studied for the biometrics of children. Fingerprint image is rich in frequency domain features therefore transform domain features should be derived.

In this paper, we are proposing a hybrid model of children fingerprint recognition by combining identification scores of DCT features with GNB classifier and CNN. As seen from Fig. 1 fingerprints of children are tiny as compared to adult fingerprints. Finding Minutia from such images is a difficult task. Finding the frequency domain features will integrate minutia and other texture features. The structure of fingerprints of two different people are shown in Fig. 2. The frequency related to each area is shown in it. The most dissimilar portion needs to separate out from the fingerprint

image. This portion is the mid to high frequency core area of the fingerprint.

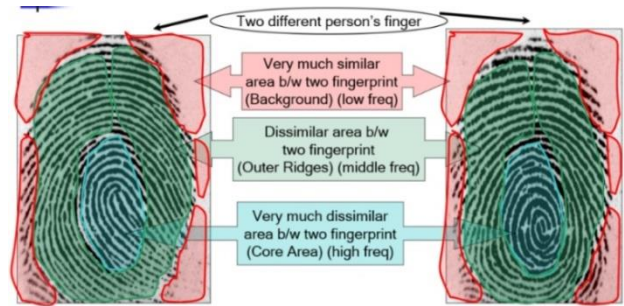


Fig. 2. Frequency distribution in fingerprint image.

The DCT has strong ability to remove correlation between the pixels like KL transform [26]. DCT allows selection of the bands directly in the frequency. Hence, we are proposing a hybrid model in which scores of classifiers GNB and CNN on the feature extracted using DCT and texture features of CNN are combined.

The contributions of research are:

- Studied transform and texture features of fingerprints for recognition of children.
- Combined Transform domain features with Machine Learning approach.
- Score level fusion of traditional method and deep learning method.
- Identification of children for this age group is studied for the first time in the literature.

The sections of the paper include Introduction, which elaborate the basic blocks and need of fingerprint recognition of children; methodology, which gives an overall idea of methods used in the proposed algorithm. Experimental Results gives information about the database used and results of different approaches. Finally, the conclusion and future scope gives insight on findings of the experimentation and future direction of study.

## II. METHODOLOGY

In this hybrid method, transform domain approaches combined with machine learning are presented and it gives promising results. The different approaches of experimentations are carried as follows:

a) DCT, Curve DCT feature extraction (Standard Deviation) and Canberra distance for feature matching is used.

b) DCT, Curve DCT used to calculate standard deviation and these features are classified using machine learning classifiers.

c) Transfer learning AlexNet model [27] used to compare with the transform domain approach.

d) The Score level fusion of Transform domain approach and Transfer learning approach is done using Max Rule, Sum Rule and Product Rule. The workflow of the research is elaborated in Fig. 3.

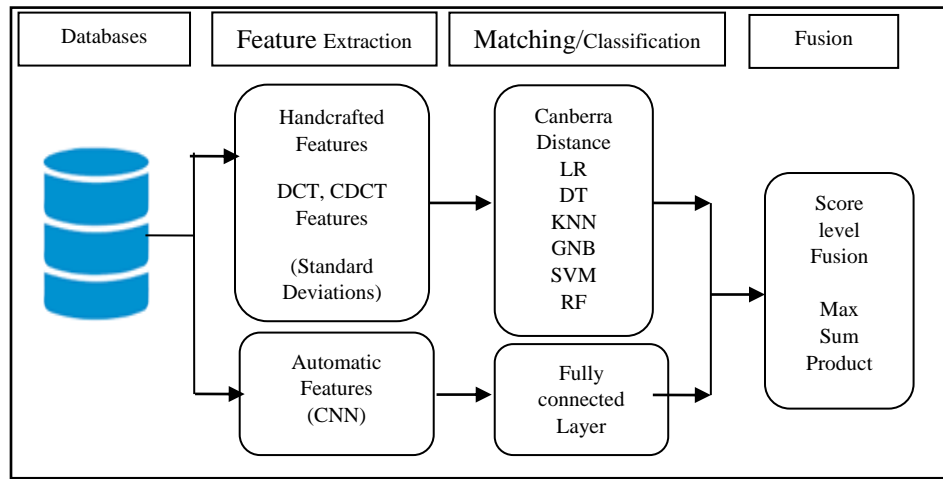


Fig. 3. Work flow of the research

A. Image Enhancement and ROI Extraction

Children's fingerprints are difficult to capture and captured fingerprints are not clear. It cannot be used directly for any algorithm. Fingerprint pre-processing and enhancement plays a crucial role in the recognition process. The most dissimilar portion shown in Fig. 2 needs to separate out from the fingerprint image. This portion is the high frequency core area of the fingerprint. To separate out the high (discriminative) frequency portion, the enhancement [28] and ROI extraction [29] steps are elaborated below.

1) *Image normalisation and segmentation*: First an input fingerprint image is normalized using mean  $M0 = 100$  and variance  $VAR0 = 100$ . The fingerprint is normalised with the help equation 1. After normalising the image segmentation is done. The standard deviation is calculated of the segmented region.

The Normalized image is defined as:

$$G(l, m) = \begin{cases} M0 + \sqrt{\frac{VAR0(I(l,m)-M)^2}{VAR}}, & \text{if } I(l, m) > M \\ M0 - \sqrt{\frac{VAR0(I(l,m)-M)^2}{VAR}}, & \text{otherwise,} \end{cases} \quad (1)$$

Where,  $I(l, m)$  are gray level values at pixel  $(l, m)$ ,  $M$  and  $VAR$  are estimated mean and variance of  $I$ ,  $G(l, m)$  is normalized Gray-level value at pixel  $(l, m)$ ,  $M0$  and  $VAR0$  are the desired mean and variance values.

2) *Local orientation estimation*: The normalized fingerprint image is used for the orientation estimation. The local ridge orientation at  $(l,m)$  is computed using eq. (2). The local ridge orientation is given in eq. (2).

$$O(l, m) = \frac{1}{2} \tan \frac{\phi'_y(l,m)}{\phi'_x(l,m)} \quad (2)$$

Where,  $\phi'_x, \phi'_y$  is image vector for lowpass filtering  $\theta(l, m)$  is the least square estimate of the local ridge orientation at the block centred at pixel  $(l, m)$ .

3) *Local frequency estimation*: The local frequency is estimated using normalized and estimated orientation image. It is given in eq. (3).

$$F(l, m) = \sum_{u=-\frac{w_l}{2}}^{\frac{w_l}{2}} \sum_{v=-\frac{w_l}{2}}^{\frac{w_l}{2}} W_l(u, v) \Omega'(1-uw), (m-vw) \quad (3)$$

Where,  $w_l$  is a two-dimensional low pass filter with unit length.  $w_l = 7$  is the size of the filter.  $\Omega'$  = interpolated frequency

4) *Mask estimation*: Normalized image is pixelwise divided into recoverable or an unrecoverable part depending upon local ridges and valleys. The three characteristics for each part are calculated, which are differences between the mean value of peak and mean value of valley. Second is the mean number of pixels between two peaks and third is the variance of the local part. These characteristics are used to classify the image in recoverable or an unrecoverable part.

5) *Filtering*: A bank of Gabor filters used as bandpass filters to remove the noise. These filters are tuned with orientation and frequency of local parts. The final enhanced image is obtained using eq. (4).

$$E(i, j) = \begin{cases} 255, & \text{if } R(i, j) = 0 \\ \sum_{u=-\frac{w_g}{2}}^{\frac{w_g}{2}} \sum_{v=-\frac{w_g}{2}}^{\frac{w_g}{2}} h(u, v : O(i, j), F(i, j)) G(i-u)(j-v), & \text{otherwise} \end{cases} \quad (4)$$

Where,  $G$  = normalized fingerprint images,  $O$  = orientation image,  $F$  = frequency image,  $R$  = recoverable mask, where  $w_g = 11$  gives the size of the Gabor filters.

6) *ROI extraction*: In the enhanced image the core point is detected using slope technique as described in paper [take from email of madam]. The original image, enhanced image and ROI extracted image is shown in Fig. 4.

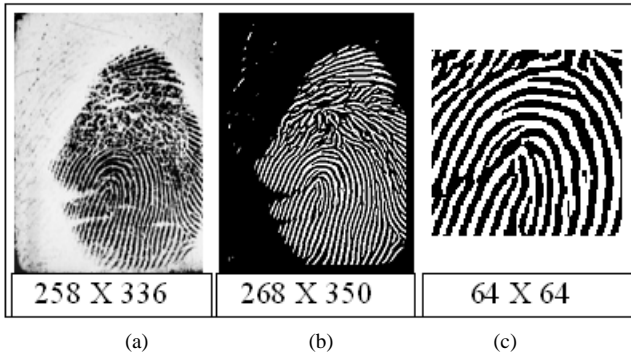


Fig. 4. (a) Original image, (b) Enhanced image, (c) ROI of enhanced image.

### B. Feature Extraction

Fingerprint texture consists of repetitive patterns of pixel intensities and it is unique to every person. To extract the texture features (standard deviations) from fingerprint images DCT, CurvedDCT are applied to the ROI of fingerprint images.

1) *DCT*: DCT is one of the most used transforms for image compression, feature extraction and recognition. The important properties of DCT are decorrelation, energy compaction and fast implementation. These properties are explored for the fingerprint feature extraction. The 2D- DCT of an image  $f(x, y)$  is given in eq. (5). The size of the image is  $M \times N$ . Variations of  $MN$  are  $U = 0$  to  $M-1$   $V = 0$  to  $N-1$ .

$$T(u, v) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cos\left(\frac{(2x+1)u\pi}{2M}\right) \cos\left(\frac{(2y+1)v\pi}{2N}\right) \quad (5)$$

The ROI is divided into four non-overlapping parts as shown in Fig. 5. The DCT is applied to each part, from each part 9 standard deviations are derived by grouping different frequency coefficients as shown in Fig. 6. Here low frequency coefficients are omitted as fingerprint image is mid and high frequency texture image as shown in Fig. 2. Total 36 features are extracted from four parts of the finger. Training is done with 4 images and average standard deviation is stored as a feature vector. Partitioning of the image gives better energy compaction and decorrelation. The mean and standard deviation is calculated from the eq. (6) and (7).

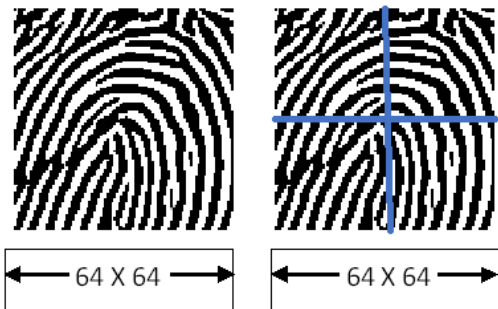


Fig. 5. (a) ROI (b) ROI is divided in four parts

The mean ( $m$ ) and standard deviation ( $sd$ ) is calculated from the eq. (6) and (7).

$$(m) = \frac{1}{N^2} \sum_{i,j=1}^N p(i, j) \quad (6)$$

$$(sd) = \sqrt{\frac{1}{N^2} \sum_{i,j=1}^N [p(i, j) - m]^2} \quad (7)$$

Thus, the total features are  $36 \times 500 = 18000$  derived from the DCT standard deviation of the images from the database.

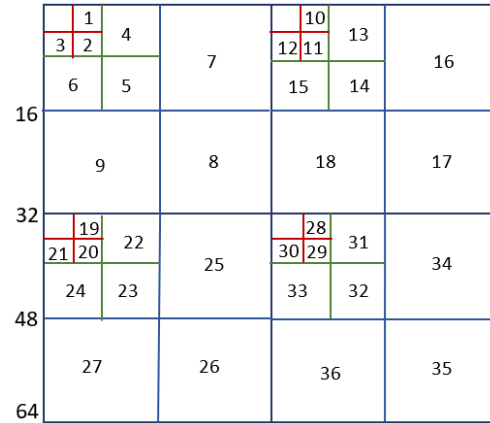


Fig. 6. DCT blocks of fingerprint image

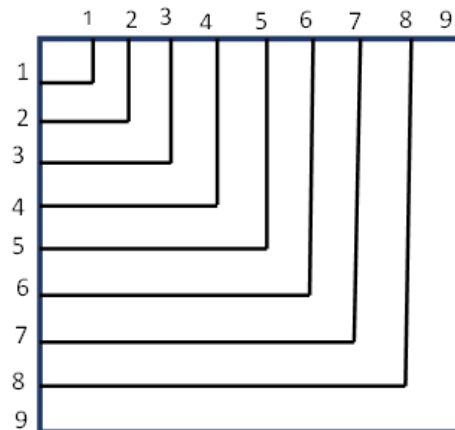


Fig. 7. Curvature bands of coefficients

### 2) Curve-DCT (CDCT)

First fingerprint image is divided as shown in Fig. 5, into four parts. For every part, 2D-DCT transform is applied and coefficients are grouped in 9 bands in a curvature manner as shown in Fig. 7. Looking at the distribution from top left corner to right bottom corner of the coefficients and their grouping, curvature form is more appropriate. For every band standard deviation is calculated. Total 36 features from four parts are stored as a training dataset.

### C. Evaluation Parameters

Evaluation of the experiments is done with class-wise identification accuracy. It is shown in eq. (8).

$$\text{Class-wise identification accuracy} = \frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i} \quad (8)$$

The  $tp_i$  is  $i^{\text{th}}$  true positive,  $tn_i$  is  $i^{\text{th}}$  true negative,  $fn_i$  is  $i^{\text{th}}$  false negative,  $fp_i$  is  $i^{\text{th}}$  false positive where  $i$  ranges from 1 to 100.

The CMC curves are used to show the Rank wise accuracy of the fusion models. Rank 1 accuracy is the accuracy achieved when the outcome is the same as ground truth at the first match by model. It can also be called as top 1 accuracy.

### III. EXPERIMENTAL RESULTS

#### A. Children Fingerprint Database

There are several fingerprint databases such as MSU-ITF, VaxTrac, CMBD [30] and NITG [24,31] of the children's fingerprints. Most of the databases are not available due to security reasons except Children's Multimodal Biometric Database (CMBD) of IIT Delhi and National Institute of Technology, Goa (NITG) database for research purposes. In this work CMBD and NITG databases are used. Total 1000 images of the left thumb from both the databases and 100 subjects from each database are used for experimentation. Both the databases used Slap fingerprint scanner with 500 ppi resolution. Some sample images of these databases are shown in Fig. 8.



Fig. 8. (a) Right thumb, (b) left thumb of CMBD [30] and (c) Right thumb (d) left thumb of NITG database [24,31]

The performance of DCT domain features is checked for children's fingerprints. The DCT is applied on cropped fingerprint images and standard deviation is calculated for a group of DCT coefficients. In the first experiment, after grouping of coefficients is done in two ways as explained in section 2.2.1 and 2.2.2. The recognition rate in identification mode is calculated using Canberra distance. This distance is normalized as seen from the eq. (8), and it is giving better accuracy than Euclidean distance. The Canberra distance between vectors a and b is in the eq. (9).

$$d(a, b) = \sum_{i=1}^n \frac{|a_i - b_i|}{|a_i + b_i|} \quad (9)$$

The recognition rate obtained is shown in Table I. The maximum recognition rate (RR) is obtained with DCT features and using a simple distance classifier for 100 subjects is 63 % for the train test split of 80-20.

TABLE I. RECOGNITION RATE (RR) WITH DCT, CDCT FEATURES AND CANBERRA DISTANCE FOR CLASSIFICATION ON CMBD AND NITG DATABASES

Features	RR for CMBD	RR for NITG
DCT	42	63
CDCT	29	37

This recognition rate (RR) is less and to improve upon it, the same feature set is given to machine learning classifiers in the next experiment. Machine learning (ML) Supervised learning classifiers are used for classification of labeled databases of children's fingerprints. These classifiers learn on their own classification of objects/images depending upon the data/features presented at the input. As machine learning are data driven models, it analyses input features, learns from it and gives the classification [32]. Choosing the best ML classifier for an application depends on the data set available, the number of features derived, data is labelled or not, variance of features etc. In this experiment the same DCT and CDCT features are given to the different classifiers such as Logistic Regression (LR), Decision Tree (DT), K- Nearest Neighbour (KNN), Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM) and Random Forest (RF). In image classification not only the features of the images but also the classification algorithms play a crucial role. The recognition rate after applying the different classifiers to the DCT, CDCT feature set is shown in Table II. As seen from the table, results obtained with CDCT features and GNB classifiers show significant improvement as compared to the first experiment. Gaussian Naïve Bayes is a Generative classifier. It requires a smaller number of parameters for classification. This classifier is probability based, where conditional and prior probabilities are calculated using Maximum Likelihood Estimation [33].

In the third experiment, the Neural Network approach is tested. As the database is small, a pre-trained AlexNet model is used for classification. Here the ROI of fingerprint image is resized to 200 X 200. Total images are split in 80-20 for training and testing. The Optimizer used is Stochastic Gradient Descent and the loss function is Sparse Categorical Cross Entropy. The recognition accuracy of AlexNet is shown in Table III for both the databases. The results obtained for CMBD database with both approaches are comparable but for NITG database DCT features and machine learning approach gives better results. Model accuracy and model loss of AlexNet approach on both the databases is shown in Fig. 9 and 10.

TABLE II. DCT AND CDCT FEATURES (STANDARD DEVIATION) EXTRACTION AND MACHINE LEARNING CLASSIFIERS ON CMBD AND NITG DATABASES

Recognition rate of CMBD database on ML classifiers						
Feature Extractor	LR	DT	KNN	GNB	SVM	RF
DCT	34.1	68.8	27.5	83.3	11.6	47
CDCT	29	71	19	92	10	44
Recognition rate of NITG database on ML classifiers						
Feature Extractor	LR	DT	KNN	GNB	SVM	RF
DCT	52.8	66.4	36.4	81.6	33.6	61
CDCT	36	66	25	96	22	45

TABLE III. COMPARISON OF PROPOSED METHOD WITH ALEXNET

Methods	NITG (100)		CMBD (100)	
	Model Accuracy	Model Loss	Model Accuracy	Model Loss
AleNet	88 %	0.5350	92 %	0.3975

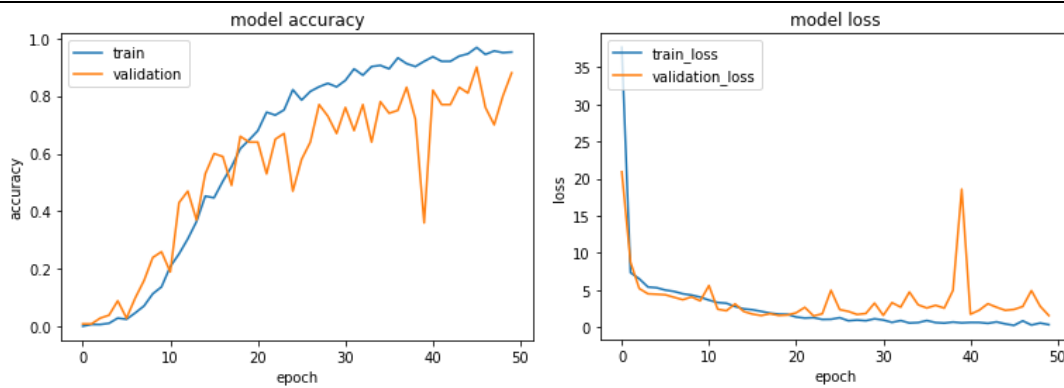


Fig. 9. AlexNet validation (a) accuracy and (b) loss on NITG database

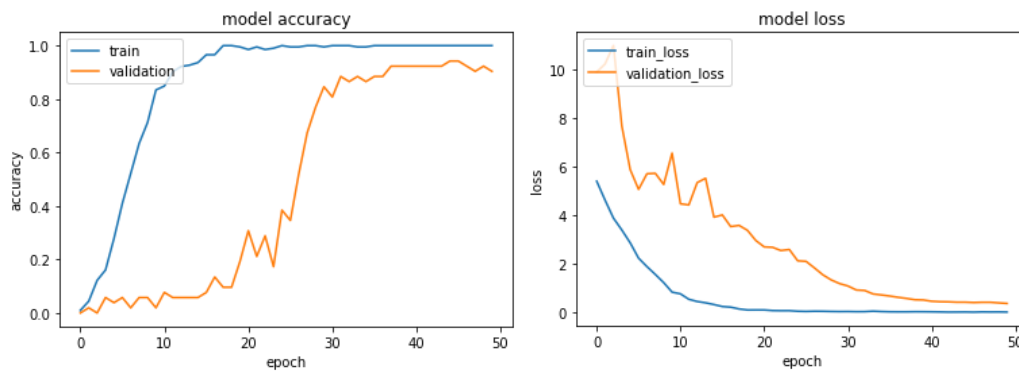


Fig. 10. AlexNet validation (a) accuracy and (b) loss on CMB database

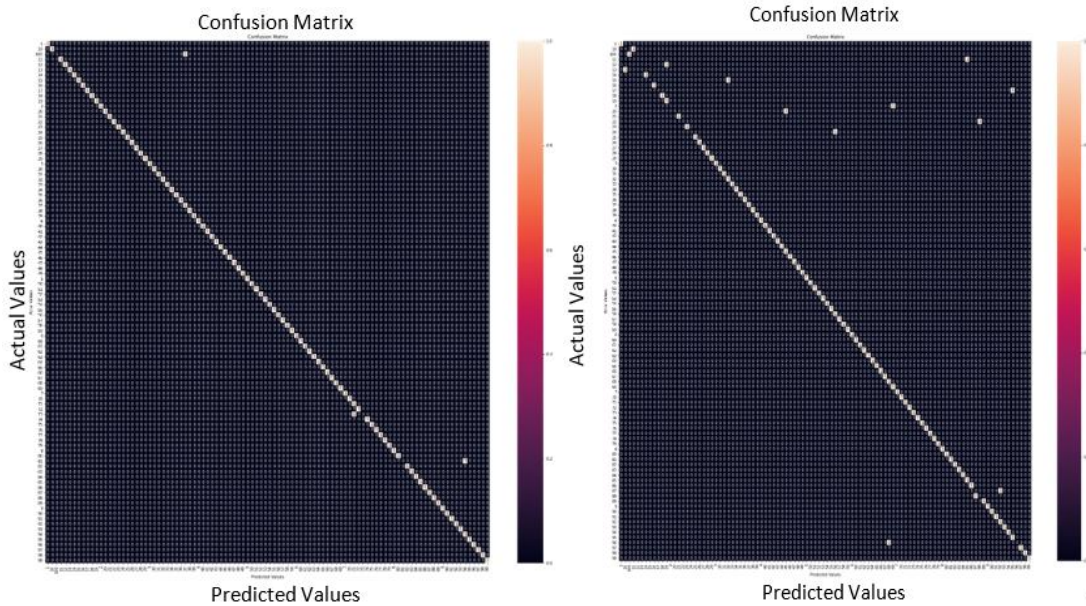


Fig. 11. Confusion matrix of AlexNet model on (a) NITG and (b) CMBD database

TABLE IV. SCORE LEVEL FUSION RESULT

Score Level Fusion	NITG		CMBD	
	Rank 1	Rank 10	Rank 1	Rank 10
Max Fusion	99	100	99	100
Sum Fusion	74.5	99	50	90
Product Fusion	56	98	43	67

The confusion matrix for both the databases is presented in the Fig. 11. In which the actual class versus predicted class is plotted. It is observed that more subjects from CMBD database are falsely classified than the NITG database as model accuracy is more for this database.

To improve the accuracy further, the hybrid mode is built by combining the scores of DCT with machine learning and AlexNet model. The Score level fusion [34] of CDCT features and Machine learning classifier with class wise identification score of AlexNet model is done. The rank1 and rank 10 accuracy for both the databases is shown in Table IV. As it is class wise identification score, max fusion is giving good accuracy because there are very few cases where both the algorithms falsely predict the same class.

The identification accuracy obtained with the proposed method is compared with the literature in Table V, Improvement in the accuracy is observed with the use of score level fusion.

TABLE V. COMPARISON OF PROPOSED METHOD WITH LITERATURE

Paper	Methods	Age	Identification Accuracy
A.K. Jain [15]	CNN, COTS	0-12 months	≤ 4 weeks 38.44 %
			> 4 weeks 73.98 %
Proposed	DCT, GNB and CNN	0 – 5 years	99 %

#### IV. CONCLUSION AND FUTURE SCOPE

This paper proposed a hybrid method of children fingerprint identification which is based on transform domain DCT and convolution feature extraction for local and global features of fingerprint images. DCT’s standard deviation features with machine learning classifier Gaussian Naïve base gives comparable accuracy as that of CNN model accuracy. We can deploy simple model based on DCT and GNB for children's fingerprint recognition, where we have fewer databases instead of the heavy CNN model. In this way, computational complexity and the need for higher version hardware is reduced. However, for higher accuracy over the period we require deep neural network models only. Here the score level combination of both algorithms is the most appropriate generalized solution for a robust model of children fingerprint identification, that is 1:N comparison is studied and implemented for children fingerprints (unimodal). The frequency domain features using DCT and CDCT, which gives compact representation of fingerprint texture image, is presented first time in this paper. The rank-1 identification accuracy achieved is 99 % using Max fusion which is higher than the literature.

The future work for this research is considering the feature level fusion. The accuracy of the children recognition can improve by using multimodal fusion with multiple modalities or multiple algorithms.

#### REFERENCES

- [1] Deb, Debayan et al. "Finding Missing Children: Aging Deep Face Features." ArXiv abs/1911.07538 (2019).
- [2] V. Kamble and M. Dale, "Face Recognition of Children Using AI Classification Approaches," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 248-251, Doi: 10.1109/ESCI50559.2021.9396891.
- [3] Debayan Deb, Neeta Nain, and Anil K. Jain. Longitudinal Study of Child Face Recognition. In 2018 International Conference on Biometrics (ICB), pages 225–232, Feb 2018.

- [4] S. Bharadwaj, H. S. Bhatt, M. Vatsa, and R. Singh. Domain specific learning for newborn face recognition. *IEEE Transactions on Information Forensics and Security*, 11(7):1630–1641, 2016.
- [5] R. P. Lemes, O. R. P. Bellon, L. Silva and A. K. Jain, "Biometric recognition of newborns: Identification using palmprints," 2011 International Joint Conference on Biometrics (IJCB), 2011, pp. 1-6, Doi:10.1109/IJCB.2011.6117475.
- [6] E. Liu, "Infant Footprint Recognition," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 1662-1669, Doi: 10.1109/ICCV.2017.183.
- [7] S. Tiwari, S. Kumar, S. Kumar and G. R. Sinha, "Ear recognition for newborn," 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), 2015, pp. 1989-1994.
- [8] Tiwari, S., & Singh, S. K. (2012), 'Newborn Verification Using Headprint', *Journal of Information Technology Research*, 5(2), 15 -30
- [9] <https://onin.com/fp/fphistory.html>, 'The History of Fingerprints' accessed on 29/09/2021.
- [10] F. Galton, "Finger Prints of Young Children," *British Association for the Advancement of Science*, vol. 69, pp. 868–869, 1899.
- [11] "Fingerprint Recognition for Children," Joint Research Center of the European Commission, Tech. Rep., 2013. [Online.] Available: <https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/fingerprint-recognition-children>.
- [12] V. Camacho et al., "Recognizing Infants and Toddlers over an On-Production Fingerprint Database," 2017 International Conference of the Biometrics Special Interest Group (BIOSIG), 2017, pp. 1-5, doi: 10.23919/BIOSIG.2017.8053518.
- [13] J. Preciozzi et al., "Fingerprint Biometrics from Newborn to Adult: A Study From a National Identity Database System," in *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2, no. 1, pp. 68-79, Jan. 2020, doi: 10.1109/TBIOM.2019.2962188.
- [14] R. Haraksim, J. Galbally, and L. Beslay, "Fingerprint growth model for mitigating the ageing effect on childrens fingerprints matching," *Pattern Recognition*, vol. 88, pp. 614–628, 2019.
- [15] Y. Koda, T. Higuchi, and A. K. Jain, "Advances in capturing child fingerprints: A high resolution CMOS image sensor with SLDR method," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2016.
- [16] Kalisky, T., Saggese, S., Zhao, Y. et al. Biometric recognition of newborns and young children for vaccinations and health care: a non-randomized prospective clinical trial. *Sci Rep* 12, 22520 (2022). <https://doi.org/10.1038/s41598-022-25986-6>.
- [17] Engelsma JJ, Deb D, Jain AK, Sudhish PS, Bhatnager A: InfantPrints: Fingerprints for Reducing Infant Mortality. arXiv e-prints. 2019.
- [18] P Macharia, P Muiruri, P Kumar, B Ngari, R Wario, "The Feasibility of Using an Android-Based Infant Fingerprint Biometrics System for Treatment Follow-Up" 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018) At: Portugal.
- [19] J. J. Engelsma, D. Deb, K. Cao, A. Bhatnagar, P. S. Sudhish and A. K. Jain, "Infant-ID: Fingerprints for Global Good," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, doi: 10.1109/TPAMI.2021.3057634.
- [20] A. K. Jain, K. Cao, and S. S. Arora, "Recognizing infants and toddlers using fingerprints: Increasing the vaccination coverage," *IEEE International Joint Conference on Biometrics*, Clearwater, FL, 2014, pp. 1-8.
- [21] Jain Anil K, Arora Sunpreet S, Best-Rowden Lacey, Cao Kai, Sudhish Prem Sewak, Bhatnagar Anjoo, Biometrics for Child Vaccination and Welfare: Persistence of Fingerprint Recognition for Infants and Toddlers. arXiv preprint arXiv:1504.04651, 2015.
- [22] A. K. Jain, S. S. Arora, L. Best-Rowden, K. Cao, P. S. Sudhish, A. Bhatnagar, and Y. Koda, "Giving infants identity: Fingerprint sensing and recognition," in *ICTD*, 2016.
- [23] A.K. Jain, A.K. Aroda, S.S. Cao, L. Best-Rowden, and A. Bhatnager, "Fingerprint recognition of young children," in *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 7, July 2017, pp. 1501-1514.
- [24] A. R. Patil, A. D. Rahulkar and C. N. Modi, "Designing an Efficient Fingerprint Recognition System for Infants and Toddlers," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-7, doi: 10.1109/ICCCNT45670.2019.8944342.
- [25] Y. Moolla, A. De Kock, G. Mabuza-Hocquet, C. S. Ntshangase, N. Nelufule and P. Khanyile, "Biometric Recognition of Infants using Fingerprint, Iris, and Ear Biometrics," in *IEEE Access*, vol. 9, pp. 38269-38286, 2021, doi: 10.1109/ACCESS.2021.3062282.
- [26] A. Jain, *Fundamentals of Digital Image Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [27] A. Krizhevsky, I. Sutskever, and G. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks", *Advances in Neural Information Processing Systems* 25, Curran Associates, Inc., (2012).
- [28] Lin Hong, Yifei Wan and A. Jain, "Fingerprint image enhancement: algorithm and performance evaluation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 777-789, Aug. 1998, doi: 10.1109/34.709565.
- [29] Anil K. Jain, Salil Prabhakar, Lin Hong and Sharad Pankanti, "Filterbank based Fingerprint Matching", *IEEE Transaction on Image Processing*, Vol. 9, No. 5, May 2000.
- [30] P. Basak, S. De, M. Agarwal, A. Malhotra, R. Singh, and M. Vatsa. Multimodal Biometric Recognition for Toddlers and Pre-School Children, In *IEEE International Joint Conference On Biometrics*, 2017.
- [31] A K Samantaray, A D Rahulkar, P J Edavoor, "A Novel Design of Dyadic db3 Orthogonal Wavelet Filter Bank for Feature Extraction", *Circuits, Systems, and Signal Processing*, Springer, Vol 40, pp.5401-5420, April 2021.
- [32] Chen, RC., Dewi, C., Huang, SW. et al. Selecting critical features for data classification based on machine learning methods. *J Big Data* 7, 52 (2020). <https://doi.org/10.1186/s40537-020-00327-4>.
- [33] Introduction to Machine Learning Notes-<https://www.cs.cmu.edu/~epxing/Class/10701-10s/Lecture/lecture5.pdf>.
- [34] Kamel Aizi, Mohamed Ouslim,, Score level fusion in multi-biometric identification based on zones of interest, *Journal of King Saud University - Computer and Information Sciences*, Volume 34, Issue 1, 2022, Pages 1498-1509, ISSN 1319-1578, <https://doi.org/10.1016/j.jksuci.2019.09.003>.



# 2-D Deep Convolutional Neural Network for Predicting the Intensity of Seismic Events

Assem Turarbek<sup>1</sup>, Yeldos Adetbekov<sup>2</sup>, Maktagali Bektemesov<sup>3</sup>

Information Systems Department, Al-Farabi Kazakh National University, Almaty, Kazakhstan<sup>1,2</sup>  
Abai Kazakh National Pedagogical University, Almaty, Kazakhstan<sup>3</sup>

**Abstract**—Machine learning has advanced rapidly in the last decade, promising to significantly change and improve the function of big data analysis in a variety of fields. When compared to traditional methods, machine learning provides significant advantages in complex problem solving, computing performance, uncertainty propagation and handling, and decision support. In this paper, we present a novel end-to-end strategy for improving the overall accuracy of earthquake detection by simultaneously improving each step of the detection pipeline. In addition, we propose a Conv2D convolutional neural network (CNN) architecture for processing seismic waveforms collected across a geophysical system. The proposed Conv2D method for earthquake detection was compared to various machine-learning approaches and state-of-the-art methods. All of the methods used were trained and tested on real data collected in Kazakhstan over the last 97 years, from 1906 to 2022. The proposed model outperformed the other models with accuracy, precision, recall, and f-score scores of 63%, 82.4%, 62.7%, and 83%, respectively. Based on the results, it is possible to conclude that the proposed Conv2D model is useful for predicting real-world earthquakes in seismic zones.

**Keywords**—Earthquake; prediction; deep learning; machine learning; classification

## I. INTRODUCTION

Over the last decade, the number and magnitude of induced earthquakes have increased dramatically, and several technological innovations [1] for effective disaster management have been developed. Nonetheless, much work remains to be done to mitigate the impact and harm caused by uncontrollable natural disasters. The proposed study is an important step toward effectively implementing modern technologies for accurate disaster detection, which remains a critical and major goal for successful emergency management.

Every year, a large number of seismic events occur around the world as a result of the release of cumulative pressure in the Earth's mantle. Catastrophic earthquakes in hilly areas may have caused several collapses on high mountainsides [2]. These and other seismic events may cause environmental issues, critical infrastructure problems, and housing developments, ultimately resulting in tragic economic losses and human casualties. Furthermore, earthquake-caused landslides dam river systems, forming lakes that may be threatened by debris flows and outburst floods that endanger people and property downstream [3, 4]. Characterizing and forecasting the geographical distribution of seismic activity landslides is recommended for disaster mitigation [5].

Determining the geographic patterns of earthquakes is difficult because the main determinants of earthquake occurrences are their parameters, landscape, soil characteristics, tectonic plates, and human impacts [6]. As a result, it may be difficult to predict where landslides will occur after an earthquake [7]. Over the last 20 years, many prediction models have been developed to pinpoint locations vulnerable to landslides caused by earthquakes, and they can be divided into two categories: (1) models with a physical and numerical foundation [8]; and (2) models for determining susceptibility [9].

The physically based models were developed using the mechanisms of gradient commencement and runoff. The first of these methods, pseudo-static analysis, proposed that the earthquake force represents an additional permanent physical force to statically conservation equations [10]. Despite the fact that selecting a pseudo-static coefficient requires criterion and always yields conservative results, pseudo-static modelling is theoretically simple [11]. Following that, the stress-deformation assessment method was proposed as an extension of finite-element simulation that is capable of simulating slope dynamic deformation.

This method, which is based on mathematical calculations and has the potential to resolve physical issues such as complicated geometries, material properties, and boundary conditions, is appropriate for investigating the stability of artificial slopes [12]. Permanent-displacement examination was proposed shortly after its deployment to calculate the displacements of landslides caused by seismic activity. Its sophistication is somewhere between the two methods discussed above. Landslides are modelled as rigid-plastic bodies sliding on an inclined plane in this analysis. The Newmark model and its variants are the most commonly used models in permanent-displacement analysis [13].

Over the last few years, advances in EQIL mechanism analysis have greatly improved the accuracy of physically based models. Because of the enormous number of parameter values required, physical-based models can only be used in a limited number of locations [11].

Later, scientists began looking into vulnerable assessment methods that could reflect a possible link between earthquake detection and causal factors for identifying earthquake-prone areas [14]. Landslide susceptibility modelling has grown in popularity over the last decade due to rapid advances in technology, geographic information systems (GIS), and data analysis [15]. These models are divided into two categories:

knowledge-driven approaches based on expert knowledge data and data-driven models based on historical landslide inventories and associated spatial landslide causative data [16].

The knowledge-based models, which use expert knowledge to explain the link between the incidence of landslides and causal causes in terms of quantity [17], believe the analytic hierarchy process to be the most representative. Numerous statistical and machine learning techniques have been developed to predict the likelihood of a landslide using data-driven models. These approaches primarily include multivariate logistic regression and artificial neural networks (ANN) [18]. According to the results, data-driven models outperform knowledge-driven models in susceptibility mapping. Data-driven models can predict earthquake dispersion patterns, which the human eye cannot [19].

The majority of the data-driven models discussed above, which are classified as classic machine-learning techniques [20-21], can represent a single layer of linear or nonlinear relationships between causal variables and the incidence of landslides. As a result, when dealing with complex data, such models are prone to overfitting or becoming trapped on a local optimum [22]. However, due to an earthquake inventory constraint that always captures an earthquake as points or polygons, these landslide prediction models rarely take the distinction between landslide source and accumulation into account.

In the last ten years, numerous deep learning algorithms have produced impressive results in computer vision, speech recognition, and intelligent robot control. Geohazard experts have gradually become aware of these algorithms' ability to exploit the potential of multiple relationships in massive data [23]. As a result, deep neural networks (DNNs), convolutional neural networks (CNNs), and their derivative models have been successfully used for landslide recognition and forecasting [24].

In seismic event prediction, there are some research gaps. The first issue in earthquake forecasting is a lack of data. Second, despite extensive research in the field of earthquake forecasting, forecasting accuracy remains low.

Third, the majority of studies employ machine learning and traditional methods that are highly dependent on the type of dataset and cannot be extrapolated to other earthquake cases.

In our research, we use the following contributions to solve these types of problems. First, we present a dataset of earthquake cases spanning the years 1906 to 2022. Second, for earthquake forecasting, we propose a Conv2D CNN model. We leave the groundwork for future research by highlighting various important parameters for training deep learning models.

Reminder of this paper is as following: Section II presents related works in the extraction and categorization of features from seismic events. Section III discusses the imagery capture procedure and demonstrates the proposed method. Section IV summarizes the findings as well as the testing and discussions surrounding the proposed approach. Section V contains a conclusion and suggestions for future research.

## II. RELATED WORKS

### A. Machine Learning in Earthquake Forecasting

Artificial intelligence techniques have been widely used to predict earthquakes [25-26]. One study [27] looked at how previous seismic occurrences in long short-term memory could be used to predict earthquake penetration rates.

Several indicators were used to determine whether seismic activity would occur within the next five minutes, including magnitude, depth, time, place, statistics, and entropy factors. Based on a spatial analysis of magnitude dispersion, an automated clustering-based adaptive neuro-fuzzy inference system for earthquake prediction was proposed [28]. However, these techniques struggle to condense useful guidelines for EQP activities [29]. As solutions to the earthquake prediction problem, several superficial machine learning experiments, such as [29-32], have been proposed. Shi et al. [33] were the first to use an artificial neural network in earthquake prediction, and they also discovered a correlation between earthquake magnitude and epicentral severity.

Subsequently, a support vector regressive and hybrid neural network was created to predict earthquakes [34]. The important indicators in this study were extracted using the criterion of greatest relevance and least redundancy. Eventually, earthquake predictions were made using traditional machine learning methods [35]. Another study [36] used a principal component analysis-based random forest to generate new datasets and reduce data dimension in order to generalize preexisting prediction models. The results show that these generalized techniques outperformed current methods in terms of average accuracy. Nonetheless, variations in geological features limit their applicability.

### B. Deep Learning in Earthquake Forecasting

Deep learning techniques are capable of calculating hundreds of complex indicators on their own. As a result, recurrent neural networks (RNNs) and convolutional neural networks (CNNs) have piqued the interest of many earthquake prediction researchers (CNNs). For example, [37] developed static-stress-based criteria for forecasting aftershock locations without assuming fault direction. It also provided a more accurate method of predicting aftershock locations and pinpointing the physical factors that governed earthquake triggering while the earthquake cycle was still active. Due to the dynamic and unpredictable nature of earthquakes, [38] developed long short-term memory to investigate the spatiotemporal association between earthquakes at various sites.

They were also able to demonstrate the dependability and efficiency of their approach. However, it is difficult for DL-based EQP techniques to produce predictions based on historical data because they require a large amount of training data to ensure accuracy [25].

Several machine-learning techniques are used on historical earthquakes to predict impending tectonic events based on earthquake waveforms. These models are used in support vector machines, random forests, k-nearest neighbours, and artificial networks [39-40]. In this study, we focus on the most

powerful RNN methods for prediction on calm and seismic days, such as LSTM models. An artificial neural network was used in one study [35] to identify earthquake precursors using TEC data, while genetic algorithms were used in another study [41].

When using machine learning algorithms to identify earthquake precursors, the TEC data of the learning pattern is considered. A-TEC data irregularity may occur prior to the earthquake in certain machine learning-based situations. In Indonesia, particularly Sumatera, efforts have been made to identify earthquake precursors using machine learning methods based on N-Model Artificial Neural Networks [42]. According to the authors of [43], QuakeCast is a one-of-a-kind technique that uses global ionosphere TEC data to identify short-term earthquakes. Using a conventional logistic regression model and a deep learning ConvLSTM autoencoder, the proposed technique investigates whether signals in an ionosphere layer TEC dataset predict earthquakes.

Without explicitly modelling specific properties, deep learning was able to forecast earthquakes. As a result, more academics are turning their attention to deep learning techniques. The authors proposed a novel ground vibration monitoring strategy for MEMS-sensed data using a deep learning approach [44]. The following study created a network for magnitude estimation using convolutional and recurrent layers [45]. In subsequent research, ConvNetQuake was developed to identify nearby micro-earthquakes based on signal waveforms. They also show how ConvNetQuake works well with different types of seismic data. Lomax et al. used CNN to quickly describe the earthquake's location, magnitude, and other characteristics [46]. S. Mostafa Mousavi investigated CNN-RNN to predict earthquakes quickly by detecting weak signals [47].

Authors then estimated the likelihood of earthquakes on the Indian subcontinent by looking at the CNN network [48]. The author in [49] investigated another CNN earthquake damage assessment model. J. A. Bayona provided two well-known seismic models to evaluate seismic dangers [50]. The experimental results indicate that certain implicit traits may be able to approach the earthquake forecasting problem from a different perspective. Although deep learning techniques can fully exploit the hidden information in earthquakes, they are not theoretically interpretable. Table I demonstrates comparison of different approaches for seismic events prediction.

TABLE I. COMPARISON OF APPROACHES FOR EARTHQUAKE MAGNITUDE PREDICTION

Reference	Applied method	Features	Dataset	Results
[33]	SVM	Magnitude	CAPCEA	69% accuracy
[34]	KMC algorithm	-	0.76	70% accuracy
[35]	HKMC, ANN	Magnitude	BMKG, USGS	Between 56% to 72% when $M \geq 6$
[36]	ACC algorithm	-	-	41.488 average distance
[37]	HKMC, ANN	-	BMKG, USGS	75% accuracy
[38]	RNN, RF, LP Boost	-	CES and USGS	79% accuracy

In this research, we want to characterize calm and earthquake days in the target station zone using total electron content (TEC) values from the ionosphere layer based on previous research in this area. The primary goal of this research is to predict faster earthquakes.

### III. DATA

It was necessary to collect data for the training sample in order to build a mathematical model. At the same time, keep in mind that the model should have constant access to new data segments in order to predict within and for a specific time period.

It is worth noting that data from the Institute of Seismology of the Republic of Kazakhstan were available during the hypothesis' development. However, because this data was only uploaded once and there was no integration with seismology institute endpoints, there was no guarantee that it could be supplied continuously. In this regard, it was decided to supplement it with additional data from the United States Geological Survey (USGS) database, which is accessible via API.

A combination of datasets from earthquake.usgs.gov and data from the Institute of Seismology of the Republic of Kazakhstan was used. Where there were columns such as place, time, magnitude, and depth. After some transformations, the dataset took the form of [year, region, rolling\_aggregations\_over\_the\_retro\_data (depth and magnitude), binary target (where 1 means there will be a devastating event, and 0 means there will not be a devastating event)].

There were 2629 events detected in the Kazakhstan area from 1906 till 2022. The final shape of the dataset after all aggregation transformations is 1170 observations and 352 parameters.

A combination of datasets from earthquake.usgs.gov and data from the Institute of Seismology of the Republic of Kazakhstan was used. Where there were columns for place, time, magnitude, and depth. After some transformations, the dataset looked like this: [year, region, rolling aggregations over the retro data (depth and magnitude), binary target (where 1 means there will be a devastating event, and 0 means there will not be a devastating event)].

From 1906 to 2022, 2629 events were detected in Kazakhstan. The dataset's final shape after all aggregation transformations is 1170 observations and 352 parameters.

The actual records span the years 1906 to 2022. There are approximately 2.5 thousand records of unique earthquake events for this period in Kazakhstan and its surroundings.

In terms of the general population, the majority of the events occurred outside of Kazakhstan (approx. 300 events). However, we believe that these events may have had an impact on Kazakhstani territory, even if they were recorded on the territory of neighboring countries in Fig. 1.

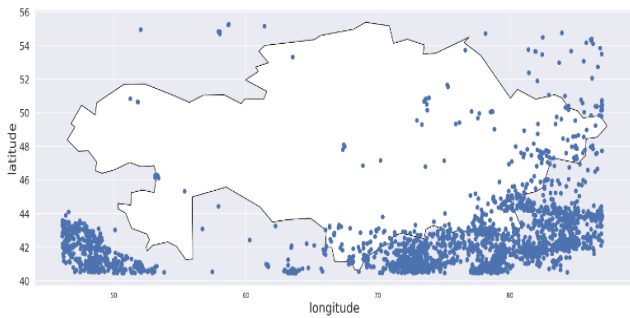


Fig. 1. Map with all events of earthquakes

Fig. 2 demonstrates distribution of earthquake cases in Kazakhstan and its neighbourhoods. Most of data are fixed after 1960. There is a peak between 1998 and 2005. Moreover, it highlights a gradual growth of earthquake cases for last 25 years, which is mainly caused by an increase of sensors and their accuracy.

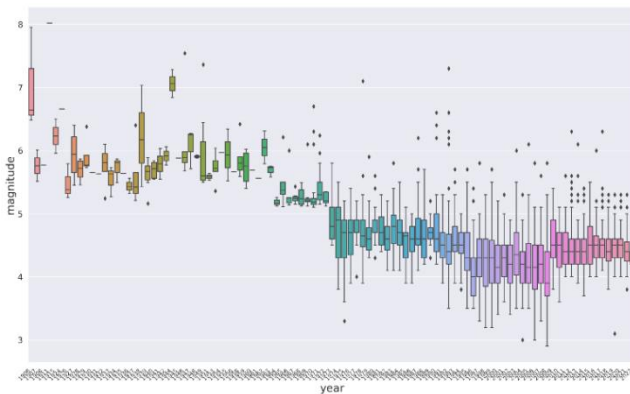


Fig. 2. Distribution of earthquakes for observed time

In order to test the hypothesis, it is advisable to divide the map of Kazakhstan into segments to predict the probability of a target event. The following methods were proposed for the segmentation of objects in the context of which predictions will be made:

- 1) Lithospheric plates
- 2) Conditional grid that divides the map of Kazakhstan (polygons)
- 3) Administrative areas

In this approach, we will try to break the map of Kazakhstan into regions, and try to aggregate the indicators grouping by regions. Since most of the events took place outside of Kazakhstan, many coordinates could not be marked with an area. Out of 2.5 thousands, 300 events remained (~11%), which is insufficient for building high-quality analytics. Nevertheless, we managed to build a baseline from this amount of data.

It is proposed to assign shocks that occurred outside the Republic of Kazakhstan, though being close to a separate area, to mark the event with a nearby area. For a purpose of forming the target events and the training sample, the main groups are specified by Area and Year parameters, aggregating the following indicators:

- 1) Minimum values in the group by the "Magnitude" and the "Depth" parameters
- 2) The maximum value in the group by the "Magnitude" and the "Depth" parameters
- 3) Median value in the group by the "Magnitude" and the "Depth" parameters
- 4) The average value in the group by the "Magnitude" and the "Depth" parameters
- 5) Standard deviation in the group by the "Magnitude" and the "Depth" parameters
- 6) Number of events in a group

Afterwards, we filled empty standard deviation values with 0. It is important to note that the sample is inconsistent over the years, as there are gaps without events between years, or they possibly were not recorded. For this reason, the following years were not indicated in the sample: 1980, 1981, 1982, 1983, 1984, 1985, and 1986.

#### IV. MATERIALS AND METHODS

In this section, we show the materials and methods used in this study. In the first section, we show how a proposed system architecture and feature extraction problem work. The following section demonstrates the proposed earthquake forecasting model. The final section shows evaluation parameters for comparing the proposed model to other machine learning models for the given problem.

##### A. Proposed System Architecture

In this research, we aimed to forecast earthquake magnitude prediction using deep learning techniques. Fig. 3 demonstrates a flowchart of our research for the prediction of earthquakes. Firstly, we get earthquake waveforms data and clean the data. The data cleaning or preprocessing process consists of four parts data cleaning, data integration, data transformation, and data reduction or dimension reduction. After preprocessing we train a deep-learning model for earthquake prediction. The architecture of the proposed deep learning model for earthquake prediction is presented in Section 3.2. The next stage is the prediction and evaluation of the proposed deep learning model.

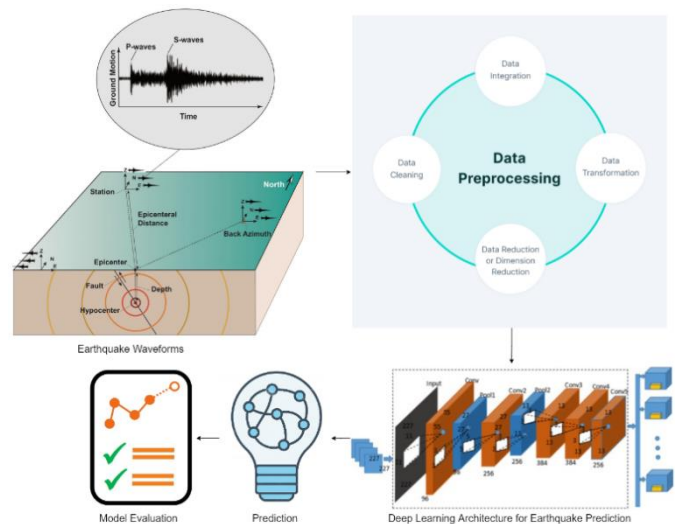


Fig. 3. Proposed architecture of earthquake forecasting

**B. Feature Extraction**

The timeline is one of the characteristics of earthquake prediction model training. Fig. 4 depicts a timeline for feature generation. Because we are at the start of the current year and have retro information about previous earthquake events (aggregated depth and magnitude parameters), our prediction is for the maximum magnitude during the year.

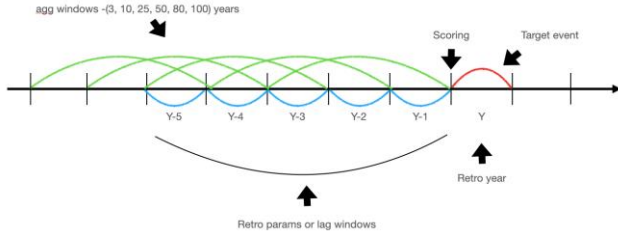


Fig. 4. Earthquake timeline as a feature

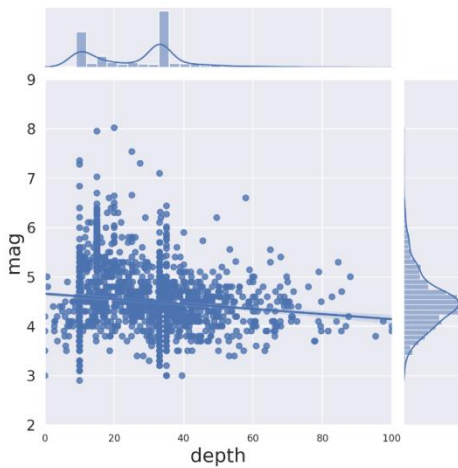


Fig. 5. Distribution of earthquakes for observed time

Fig. 5 depicts a relationship between depth and magnitude. While smaller earthquakes can and do occur at all depths down to around 700 km, the largest earthquakes occur at shallower depths in the earth's crust. Earthquakes occur in the crust, the earth's highest layer, which ranges in thickness from 7 to 30 km. The earth's crust, which contains numerous fault networks that can cause earthquakes, is the planet's coldest and most vulnerable region. These earthquakes are caused by frictional sliding on faults caused by tectonic stress accumulation.

**C. Proposed Model**

The proposed deep learning prediction architecture employs the Convolutional neural network strategy, as illustrated in Fig. 6. A Rectified Linear Unit (ReLU) activation function layer precedes the Maxpooling2D layer with a filter size of (33) and is followed by a Conv2D layer with a size of (128) and a filter size of (33). The first epoch assigns the autoencoder model's received output to these layers. The second stage, like the first, employs a Conv2D layer with a size of (64). From the third to the eighth stage, only the Conv2D layer with a filter size of (33) and ReLU activation are operational.

The output of the eighth stage is flattened by the ninth stage. Following that, in stages 10 and 11, we use completely linked layers with 50 and 10 neurons, respectively. Two scenarios are also included in the proposed regression model. Initially, we use a single output neuron to estimate the magnitude. Second, we use three neurons to estimate both magnitude and position.

The important feature mappings, which are retrieved by each convolutional layer in the proposed CNN technique, are adapted as a matrix of pixels from an image. Each feature mapping is identified by equation (1):

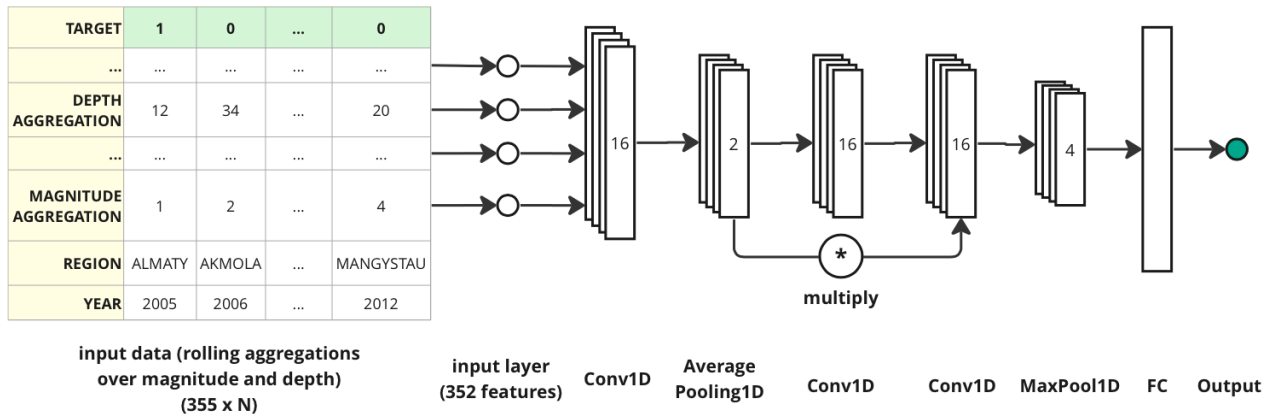


Fig. 6. Proposed Conv2D CNN architecture

$$\begin{aligned}
 FM[m, n] &= (g * k)[m, n] \\
 &= \sum_i^{\phi} \sum_j^{\psi} (X^{CNN}[m + i, n + j] * C[i, j]) \quad (1) \\
 \forall m, n \notin X^{CNN}, m \neq n,
 \end{aligned}$$

The critical feature mappings are extracted from an image as a matrix of pixels by each convolutional layer in the proposed Conv 2D convolutional neural network approach. Equation (2) identifies each feature mapping of the proposed approach.

$$f(m, n) = \max(0, FM(m, n)) \quad (2)$$

The Maxpooling layer, which is described by following formula is subsequently used to increase the number of feature maps, deepen the proposed neural network, and reduce the network dimension. Equation (3) describes maxpooling layer for the given deep learning model.

$$MP(m, n) = \max f(m + l, n + d) \quad (3)$$

where l and d are the Maxpooling window dimension.

#### D. Evaluation Method

The prediction outcomes are analyzed using the metrics of accuracy, precision, recall, and F1-score [43-46]. The accuracy indicator displays the rate of model prediction accuracy across all parameters. It is calculated as the proportion of correct predictions made by a model. This is especially useful when all of the courses have the same value. It is calculated by dividing the number of correct predictions by the total number of predictions made. This is the probability that the class will be adequately anticipated. The precision of the formula is shown in Equation (4).

$$Accuracy(a) = \frac{\sum_{i=1}^N [a(x_i) = y_i]}{N} \quad (4)$$

$$= \frac{TP + TN}{TP + TN + FP + FN}$$

Here, TP is true positives, TN is true negatives, FP is false positives, FN is false negatives. Sum of all this cases gives a number of all cases.

Precision offers a reliable picture of the veracity of our positive detections in comparison to the unchanging truth. How many of the objects we predicted in a given image matched the ground truth annotation? Formula (5) describes precision [47].

$$precision = \frac{TP}{TP + FP} \quad (5)$$

When striving to accurately describe the extent to which our pessimistic expectations match the reality, recall or sensitivity is a good statistic to utilize. Out of all the challenges in our ground truth, how many favorable forecasts did we get? [47]

$$recall = \frac{TP}{TP + FN} \quad (6)$$

The symbol F-measure stands for the harmonic mean of accuracy and completeness. This statistic decreases as accuracy or completeness approach zero. The formula for the F-measure evaluation parameter is shown in equation (7).

$$F_{measure} = \frac{2 Precision Recall}{Precision + Recall} \quad (7)$$

#### V. EXPERIMENT RESULTS

This section presents the experimental results of the proposed model for earthquake detection and compares them to

classical machine learning methods and state-of-the-art models. The confusion matrix for the given problem is shown in Fig. 4. The proposed deep model has a high prediction percentage, as evidenced by the results.

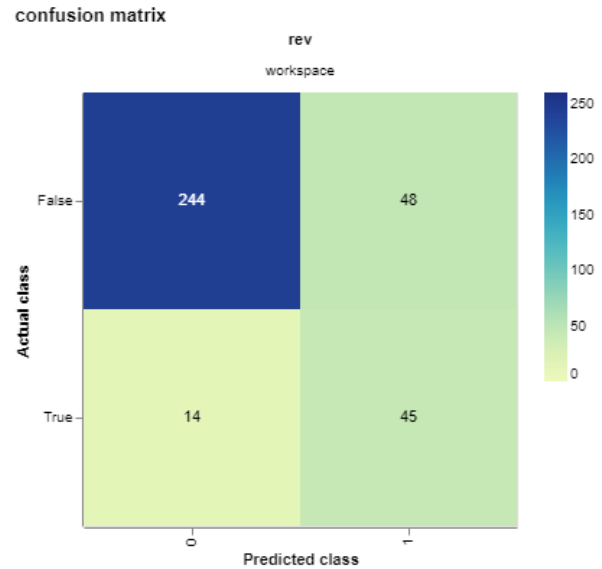


Fig. 7. Earthquake timeline as a feature

Fig. 7 depicts the results of earthquake forecasting. There, we show the results of three machine learning algorithms for earthquake detection after ten training epochs. As the results show, the light gradient boosting machine (lightgun) outperforms other machine learning methods in terms of accuracy and ROC-AUC. In each evaluation parameter, the neural network has the lowest efficiency. Random Forest performs well in some parameters such as ROC-OUC, recall.

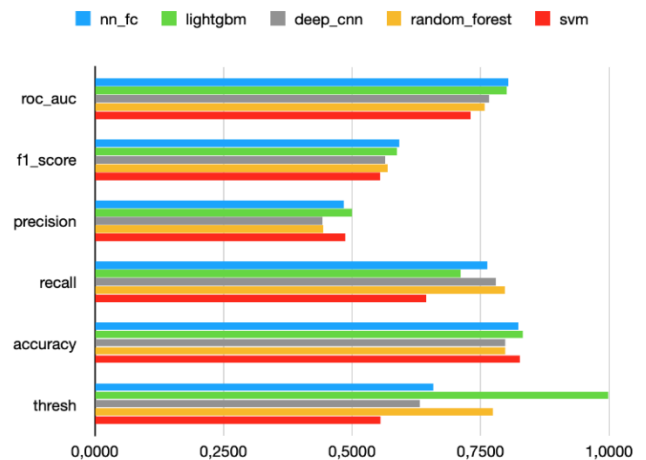


Fig. 8. Precision and recall for 10 epochs

Table II compares various machine learning algorithms and the proposed deep learning model for earthquake forecasting problems. According to the test results, the proposed deep model outperforms traditional machine learning algorithms in every evaluation parameter. It means that the proposed deep model is practical (see Fig. 8).

TABLE II. COMPARISON OF APPROACHES FOR EARTHQUAKE MAGNITUDE PREDICTION

Algorithm	Accuracy	Precision	Recall	F-score	AUC-ROC	Threshold
Proposed Model	0.874	0.63	0.824	0.627	0.83	0.998
LightGBM	0.832	0.500	0.712	0.587	0.801	0.998
Random Forest	0.795	0.443	0.797	0.570	0.758	0.774
Neural Network	0.766	0.365	0.525	0.431	0.595	0.766
SVM	0.742	0.387	0.506	0.428	0.587	0.628
Decision Tree	0.512	0.536	0.483	0.417	0.561	0.637

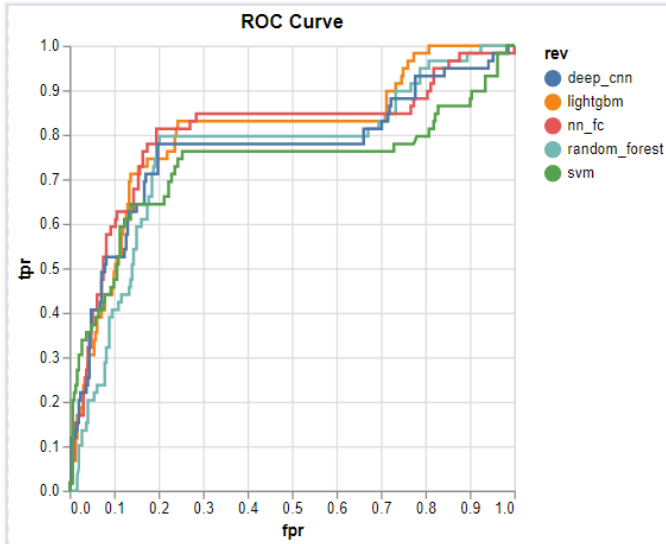


Fig. 9. ROC curve for 10 epochs

The area under the curve receiver operating characteristics (AUC-ROC) curve for the proposed model for earthquake forecasting problems is depicted in Fig. 9 in terms of the combination of false positive rate and true positive rate for 10 training epochs. The AUC-ROC curve is greater than 0.5, indicating that the proposed deep model is practically acceptable and feasible in real life. The horizontal axis stands for false positive rates, the vertical axis means true positive rate. The results show that the proposed deep convolutional neural network demonstrates high AUC-ROC value by achieving higher value than the other applied models in ten epochs. Obtained results show that the proposed model can be applied for real case (see Fig. 10).

## VI. DISCUSSION AND FUTURE RESEARCH

As shown in the figures above, the results appear promising, though it is worth noting that the model is not stable due to a lack of training data. The hypothesis is that earthquake events occur in cycles, and that retro data based solely on magnitude and depth predictors can predict the future appearance of destructive earthquakes. Machine learning algorithms are clearly based solely on statistics, and they should not be regarded as magical black boxes. The model has no idea what an earthquake is. There is a hypothesis that goes something like this: "In order for computers to understand earthquake concepts, we should pass fundamental features that describe physical concepts of earthquake nature."

Many suggestions are aimed at improving the model. For example, a hypothesis is proposed to test with new features such as: The impact of lithospheric plate movements:

- The impact of natural disasters and macroeconomic indicators on climate change
- The impact of events in neighboring regions (Relative features)
- Mining's impact on deposits
- Formulas for physical sense.

Other suggestions include: Scaling down the grid - Making predictions in quarters and months rather than years.

- Create a project with a trigger so that local emergency services can respond.

## VII. CONCLUSION

The proposed approach demonstrates that the presented deep learning model outperforms approaches using traditional

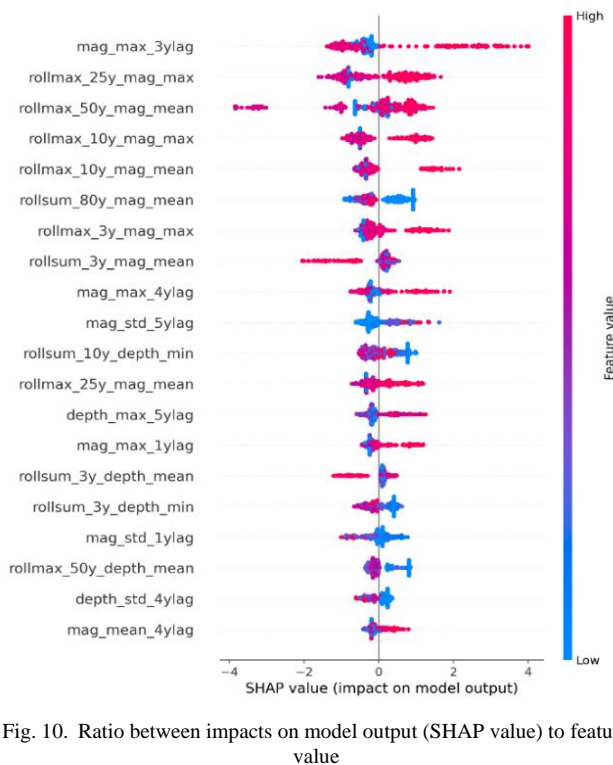


Fig. 10. Ratio between impacts on model output (SHAP value) to feature value

machine learning methods and cutting-edge deep models that employ traditional features. The provided time series-based method has the potential to improve the accuracy of earthquake forecasting issues. The applied dataset of earthquakes for the years 1906 to 2022 with magnitudes between 4 and 7 increases by 8.5% using the proposed deep learning architecture with the provided features, indicating that the proposed approach is somewhat successful on datasets with a pretty large size. Recent studies have shown that massive data analytics and machine learning can improve earthquake prediction accuracy. The proposed deep learning model performed well in earthquake forecasting, with accuracy, precision, recall, f-score, and AUC-ROC of 87.4%, 63%, 82.4%, 62.7%, and 83%, respectively. Specifically, incorporating the proposed deep learning architecture provided spatial and temporal characteristics, allowing earthquakes to be predicted to some extent.

#### REFERENCES

- [1] W. Zhu and G. Beroza, "PhaseNet: a deep-neural-network-based seismic arrival-time picking method," *Geophysical Journal International*, vol. 216, no. 1, pp. 261–273, 2019.
- [2] X. Fan, G. Scaringi, O. Korup, A. West, C. van Westen, H. Tanyas, R. Huang, "Earthquake-induced chains of geologic hazards: Patterns, mechanisms, and impacts," *Reviews of Geophysics*, vol. 57, no. 2, pp. 421–503, 2019.
- [3] P. Alfaro, J. Delgado, F. García-Tortosa, L. Lenti, J. López, C. López-Casardo, S. Martino, "Widespread landslides induced by the Mw 5.1 earthquake of 11 May 2011 in Lorca, SE Spain," *Engineering Geology*, vol. 137, pp. 40–52, 2012.
- [4] P. Cui, Y. Zhu, Y. Han, X. Chen and J. Zhuang, "The 12 May Wenchuan earthquake-induced landslide lakes: distribution and preliminary risk evaluation," *Landslides*, vol. 6, no. 3, pp. 209–223, 2009.
- [5] C. Ye, Y. Li, P. Cui, L. Liang, S. Pirasteh, J. Marcató, J. Li, "Landslide detection of hyperspectral remote sensing data based on deep learning with constraints," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, pp. 5047–5060, 2019.
- [6] P. Hernández, J. Ramírez and M. Soto, "Deep-learning-based earthquake detection for fiber-optic distributed acoustic sensing," *Journal of Lightwave Technology*, vol. 40, no. 8, pp. 2639–2650, 2022.
- [7] O. Korup and A. Stolle, "Landslide prediction from machine learning," *Geology Today*, vol. 30, no. 1, pp. 26–33, 2014.
- [8] A. Salem, M. Yakoot and O. Mahmoud, "Numerical-Based Model for Calculating the Risk of Well Integrity Failures in Mature Fields Operated by Gas Lift," *Journal of Failure Analysis and Prevention*, vol. 22, no. 2, pp. 757–772, 2014, 2022.
- [9] A. Mosavi, M. Golshan, S. Janizadeh, B. Choubin, A. Melesse, A. Dineva, "Ensemble models of GLM, FDA, MARS, and RF for flood and erosion susceptibility mapping: a priority assessment of sub-basins," *Geocarto International*, vol. 37, no. 9, pp. 2541–2560, 2022.
- [10] S. Muthukumar, S. Kolathayar, A. Valli and D. Sathyan, "Pseudostatic analysis of soil nailed vertical wall for composite failure," *Geomechanics and Geoengineering*, vol. 17, no. 2, pp. 561–573, 2022.
- [11] J. Bojadjieva, V. Sheshov, and C. Bonnard, "Hazard and risk assessment of earthquake-induced landslides—case study," *Landslides*, vol. 15, no. 1, pp. 161–171, 2018.
- [12] G. Rios, J. Sebastián Rincón Tabares, A. Montoya, D. Restrepo and H. Millwater, "Transient thermomechanical sensitivity analysis using a complex-variable finite element method," *Journal of Thermal Stresses*, vol. 45, no. 5, pp. 341–374, 2022.
- [13] P. Nayek and M. Gade, "Artificial neural network-based fully data-driven models for prediction of newmark sliding displacement of slopes," *Neural Computing and Applications*, vol. 34, no. 11, pp. 9191–9203, 2022.
- [14] H. Wen, X. Wu, S. Ling, C. Sun, Q. Liu, G. Zhou, "Characteristics and susceptibility assessment of the earthquake-triggered landslides in moderate-minor earthquake prone areas at southern margin of Sichuan Basin, China," *Bulletin of Engineering Geology and the Environment*, vol. 81, no. 9, pp. 1–20, 2022.
- [15] Y. Yi, Z. Zhang, W. Zhang, Q. Xu, C. Deng, Q. Li, "GIS-based earthquake-triggered-landslide susceptibility mapping with an integrated weighted index model in Jiuzhaigou region of Sichuan Province," *China Natural Hazards and Earth System Sciences*, vol. 19, no. 9, pp. 1973–1988, 2019.
- [16] F. Huang, J. Zhang, C. Zhou, Y. Wang, J. Huang, L. Zhu, "A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction," *Landslides*, vol. 17, no. 1, pp. 217–229, 2020.
- [17] S. Das, S. Sarkar and D. Kanungo, "GIS-based landslide susceptibility zonation mapping using the analytic hierarchy process (AHP) method in parts of Kalimpong Region of Darjeeling Himalaya," *Environmental Monitoring and Assessment*, vol. 194, no. 3, pp. 1–28, 2022.
- [18] M. Maya, W. Yu and L. Telesca, "Multi-step forecasting of earthquake magnitude using meta-learning based neural networks," *Cybernetics and Systems*, vol. 53, no. 6, pp. 563–580, 2022.
- [19] K. Saini, S. Kalra and S. Sood, "An Integrated Framework for Smart Earthquake Prediction: IoT, Fog, and Cloud Computing," *Journal of Grid Computing*, vol. 20, no. 2, pp. 1–20, 2022.
- [20] B. Omarov, N. Saparkhojayev, S. Shekerbekova, O. Akhmetova, M. Sakypbekova, G. Kamalova, Zh. Alinzhanova, L. Tukenova, Zh. Akanova, "Artificial intelligence in medicine: real time electronic stethoscope for heart diseases detection," *Computers, Materials & Continua*, vol. 70, no. 2, pp. 2815–2833, 2022.
- [21] B. Omarov, A. Tursynova, O. Postolache, K. Gamry, A. Bатыrbekov, S. Aldeshov, Zh. Azhibekova, M. Nurtas, A. Aliyeva, K. Shiyapov, "Modified unet model for brain stroke lesion segmentation on computed tomography images," *Computers, Materials & Continua*, vol. 71, no. 3, pp. 4701–4717, 2022.
- [22] G. Yao, J. Wang, B. Cui and Y. Ma, "Quantifying effects of tasks on group performance in social learning," *Humanities and Social Sciences Communications*, vol. 9, no. 1, pp. 1–11, 2022.
- [23] B. Omarov, A. Bатыrbekov, K. Dalbekova, G. Abdulkarimova, S. Berkimbaeva, S. Kenzhegulova, F. Gusmanova, "Electronic stethoscope for heartbeat abnormality detection," in *5th Int. Conf. on Smart Computing and Communication (SmartCom 2020)*, Paris, France, pp. 248–258, 2020.
- [24] A. Altayeva, B. Omarov and Y. Cho, "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)*, Shanghai, China, pp. 693–696, 2018.
- [25] M. Al Banna, K. Taher, M. Kaiser, M. Mahmud, M. Rahman, A. Hosen, G. Cho, "Application of artificial intelligence in predicting earthquakes: state-of-the-art and future challenges," *IEEE Access*, vol. 8, no. 1, pp. 192880–192923, 2020.
- [26] M. Al-Rawashdeh, I. Yousef and M. Al-Nawaiseh, "Predicting the Inelastic Response of Base Isolated Structures Utilizing Regression Analysis and Artificial Neural Network," *Civil Engineering Journal*, vol. 8, no. 6, pp. 1178–1193, 2022.
- [27] B. Gao, R. Wang, C. Lin, X. Guo, B. Liu, W. Zhang, "TBM penetration rate prediction based on the long short-term memory neural network," *Underground Space*, vol. 6, no. 6, pp. 718–731, 2021.
- [28] C. Jiang, X. Wei, X. Cui and D. You, "Application of support vector machine to synthetic earthquake prediction," *Earthquake Science*, vol. 22, no. 3, pp. 315–320, Jun. 2009.
- [29] R. Kamath and R. Kamat, "Earthquake magnitude prediction for andaman-nicobar Islands: adaptive neuro fuzzy modeling with fuzzy subtractive clustering approach," *Journal of Chemical and Pharmaceutical Sciences*, vol. 10, no. 3, pp. 1228–1233, 2017.
- [30] Anand, M., Sahay, K. B., Ahmed, M. A., Sultan, D., Chandan, R. R., & Singh, B. (2022). Deep learning and natural language processing in computation for offensive language detection in online social networks by feature selection and ensemble classification techniques. *Theoretical Computer Science*.



- [31] Sultan, D., Toktarova, A., Zhumadillayeva, A., Aldeshov, S., Mussiraliyeva, S., Beissenova, G., Tursynbayev, A., Baenova, G. & Imanbayeva, A. (2023). Cyberbullying-related hate speech detection using shallow-to-deep learning. *CMC-COMPUTERS MATERIALS & CONTINUA*, 74(1), 2115-2131.
- [32] H. Adeli and A. Panakkat, "A probabilistic neural network for earthquake magnitude prediction," *Neural Networks*, vol. 22, no. 7, pp. 1018–1024, 2009.
- [33] C. Shi and X. Liu, "Application of neural network to earthquake engineering," *Earthquake Engineering and Engineering Vibration*, vol. 11, no. 2, pp. 39–46, 1991.
- [34] M. Shodiq, D. Kusuma, M. Rifqi, A. Barakbah and T. Harsono, "Neural network for earthquake prediction based on automatic clustering in indonesia," *International Journal on Informatics Visualization*, vol. 2, no. 1, pp. 37–43, 2018.
- [35] M. Shodiq, D. Kusuma, M. Rifqi, A. Barakbah and T. Harsono, "Spatial analysis of magnitude distribution for earthquake prediction using neural network based on automatic clustering in Indonesia," in *Proc. International Electronics Symposium on Knowledge Creation and Intelligent Computing*, Surabaya, Indonesia, pp. 246–251, 2017.
- [36] X. Shao, X. Li, L. Li and X. Hu, "The application of ant-colony clustering algorithm to earthquake prediction," in *Advances in Electronic Engineering, Communication and Management*, vol. 2, no. 1, pp. 145–150, 2012.
- [37] P. DeVries, F. Viégas, M. Wattenberg and B. Meade, "Deep learning of aftershock patterns following large earthquakes," *Nature*, vol. 560, no. 7720, pp. 632–634, 2018.
- [38] Q. Wang, Y. Guo, L. Yu and P. Li, "Earthquake prediction based on spatio-temporal data mining: an LSTM network approach," *IEEE Transactions on Emerging Topics in Computing*, vol. 8, no. 1, pp.148–158, 2017.
- [39] J. Mahmoudi, M. Arjomand, M. Rezaei and M. Mohammadi, "Predicting the earthquake magnitude using the multilayer perceptron neural network with two hidden layers", *Civil Engineering Journal*, vol. 2, no. 1, pp. 1–12, 2016.
- [40] K. Asim, F. Martínez-Álvarez, A. Basit and T. Iqbal, "Earthquake magnitude prediction in hindukush region using machine learning techniques," *Natural Hazards*, vol. 85, no. 1, pp. 471–486, 2017.
- [41] M. Akhoondzadeh, "Genetic algorithm for TEC seismo-ionospheric anomalies detection around the time of the Solomon (Mw= 8.0) earthquake of 06 February 2013", *Advances in Space Research*, vol. 52, no. 4, pp. 581–590, 2013.
- [42] B. Aji, T. Liong and B. Muslim, "Detection precursor of sumatra earthquake based on ionospheric total electron content anomalies using N-Model Artificial Neural Network," in *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, IEEE. pp. 269–276, 2017.
- [43] A. Rayan and H. Artuner, "LSTM-Based Deep Learning Methods for Prediction of Earthquakes Using Ionospheric Data," *Gazi University Journal of Science*, vol. 35, no. 4, pp. 1417–1431, 2022.
- [44] I. Habibagahi, M. Omidbeigi, J. Hadaya, H. Lyu, J. Jang, J. Ardell, A. Babakhani, "Vagus nerve stimulation using a miniaturized wirelessly powered stimulator in pigs," *Scientific Reports*, vol. 12, no. 1, pp. 1–12, 2022.
- [45] V. Ocegueda-Hernández, I. Román-Godínez and G. Mendizabal-Ruiz, "A lightweight convolutional neural network for pose estimation of a planar model," *Machine Vision and Applications*, vol. 33, no. 3, pp. 1–21, 2022.
- [46] A. Lomax, A. Michelini and D. Jozinović, "An investigation of rapid earthquake characterization using single - station waveforms and a convolutional neural network," *Seismological Research Letters*, vol. 90, no. 2A, pp. 517–529, 2019.
- [47] S. Mousavi, W. Zhu, Y. Sheng and G. Beroza, "CRED: a deep residual network of convolutional and recurrent units for earthquake signal detection," *Scientific Reports*, vol. 9, no. 1, pp. 1–14, 2019.
- [48] J. Núñez, P. Catalán, C. Valle, N. Zamora and A. Valderrama, "Discriminating the occurrence of inundation in tsunami early warning with one-dimensional convolutional neural networks," *Scientific Reports*, vol. 12, no. 1, pp. 1–20, 2022.
- [49] R. Ünlü and R. Kiriş, "Detection of damaged buildings after an earthquake with convolutional neural networks in conjunction with image segmentation," *The Visual Computer*, vol. 38, no. 2, pp. 685–694, 2022.
- [50] J. Bayona, W. Savran, A. Strader, S. Hainzl, F. Cotton, D. Schorlemer, "Two global ensemble seismicity models obtained from the combination of interseismic strain measurements and earthquake-catalogue information," *Geophysical Journal International*, vol. 224, no. 3, pp. 1945–1955, 2021.

# User-Centered Design (UCD) of Time-Critical Weather Alert Application

Abdulelah M. Ali<sup>1</sup>, Abdulrahman Khamaj<sup>2</sup>, Ziho Kang<sup>3</sup>, Majed Moosa<sup>4</sup>, Mohd Mukhtar Alam<sup>5</sup>

Department of Industrial Engineering, Jazan University, Jazan, Saudi Arabia<sup>1,2,4</sup>

School of Industrial and Systems Engineering-College of Engineering, University of Oklahoma, Oklahoma<sup>4</sup>

Department of Mechanical Engineering, Vivekananda Global University, Jaipur, Rajasthan, India<sup>5</sup>

**Abstract**—Weather alert applications can save precious lives in time-critical risk situations; however, even the most widely used applications may fall short in intuitive interface and content design, possibly due to limitations in the users participation in the design process and in the users range considered. The objective of this study was to investigate whether the application of UCD principles and usability guidelines can improve the use of and satisfaction of time-critical weather alert apps by the public and or expert users. A prototype of a UCD-based weather alert application was developed and evaluated. Initially, thirty-two voluntaries participated in the identification of the important features that lead to the development of the porotype, and then the prototype was tested with another eighty participants (40 young and 40 elderly). The prototype includes five enhancements: auto-suggested location search, an all-inclusive interface for weather forecasts, message alert, visual and intuitive map settings, and minimalism-oriented alert settings. The enhanced functionality was compared to similar functionality in existing commercial weather applications. Effectiveness (completion rate, error count, error severity, and error cause), efficiency (time to completion), and satisfaction (post-task and post-test surveys) were measured. The results showed the enhancements significantly improved performance and satisfaction across both age groups compared to equivalent functionality in the existing app. The Mann-Whitney U test showed a statistically significant difference ( $p < 0.001$ ) in task satisfaction and number of errors between the two apps for all tasks. The Mann-Whitney U test showed a significant difference ( $p < 0.001$ ) in the across all tasks between the two apps Also, overall, young people with existing apps outperformed elderly, and both young and elderly with enhanced apps performed very high. Therefore, the enhancements implemented through the UCD process and usability guidelines significantly improved performance and satisfaction across both age groups to facilitate timely action necessary during a crisis.

**Keywords**—User-centered design; time-critical weather alert apps; weather forecasts; map set-tings; message alert

## I. INTRODUCTION

With the advancement of smartphone technology, countless people have started to receive weather information on the smartphone platform. Due to the requirement for immediate action weather apps are important and potentially life-saving tools in time-critical severe weather events such as flooding and tornadoes. Specifically, usability issues with a poorly designed weather app interface can disproportionately affect certain end-user groups, especially the elderly. Elder users often experience problems with age, such as decreased working

memory, and decreased visual, motor abilities, difficulties and cognitive [1-3]. In addition to physical and cognitive difficulties, elderly often struggle to cope with the rapid development of smartphone technology [4]; however, smartphone usage among older users continues to grow. As of 2017, nearly 74% of the U.S. population aged 50 to 64 and 42% of the population aged 65 and older owned a smartphone [5]. Numerous studies have focused on age differences in the use of smartphone apps in various fields, including information technology [6], healthcare [7], and communication [8]. However, to our knowledge, there are limited studies on usability testing of weather apps [9, 10], and no one has applied user-centered design (UCD) to include a wide range of users, especially elderly.

Drogalis et al. [10] identified usability issues with weather alert applications, such as the inclusion of hidden map menus that required prior counterintuitive actions and the lack of feedback on actions that were performed. Khamaj and Kang [9] investigate usability issues such as poor visualization of critical weather information, inappropriate language usage in time-critical alert messages, and inefficient location search functionality. However, neither of the two studies mentioned above addresses the question of how to effectively meet users' issues and needs. One possible way to address usability issues is to apply a user-centered design (UCD) approach, especially when it comes to end-users of different age groups. In detail, UCD refers to a cycle of design stages in which developers consider the needs, capabilities, and constraints of target users at each stage [11]. Norman [12] and Mao et al. [13] established a general UCD framework and applied to the human factors by Schnall et al. [14]. Careful consideration of each stage is believed to result in an interface that is easy to use and deemed useful [15]. The purpose of this study is to investigate whether the UCD approach could address usability issues for young and elderly when interacting with a smartphone weather alert application.

## II. METHOD

The specific contribution of this research is to (1) characterize and classify the issues and needs of users with a broad age range in the design of weather alert applications, and (2) relate these issues and needs to usability guidelines (or principles), (3) design a prototype based on these guidelines, and (4) evaluate the prototype by selecting a target group (i.e., young and elderly). The results could impact the entire design

process of a weather alert app (or other similar apps) that could save lives in a time-critical event.

Fig. 1 provides a summary of the framework used and applied by the afore-mentioned researchers. Following the framework provided in Fig. 1, the first stage of the UCD process is to describe the user's issues and needs when interacting with an existing application. Issues and requirements can be aligned with major usability guidelines, such as those presented by Nielsen [16] and Rogers et al. [17] and smartphone app design specifications, such as those presented by Gove [18]. At this stage, a broad range of users can be recruited to investigate various problems and needs that they can characterize and classified.

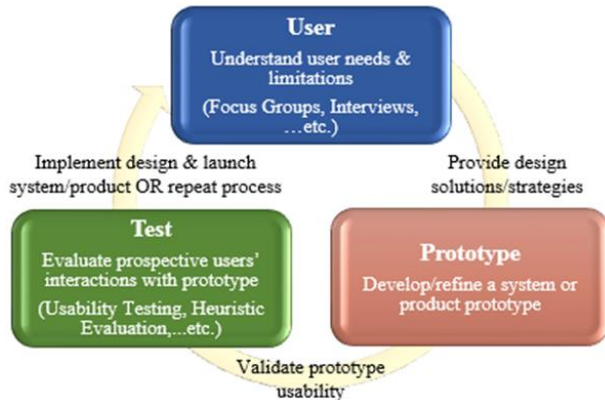


Fig. 1. Summary of the user-centered design framework (Jung et al. [15])

Specific to the weather app, questions and needs may relate to any of the features available on any typical weather app, such as location search, weather forecast, alert messages/notifications, alert settings, and radar/maps. The analyzed data can be characterized and classified using content analysis methods. One possible approach is thematic analysis [19]. This approach has been applied to critically evaluate and describe datasets, as well as to implicitly and explicitly identify provided ideas in form of codes refined as themes and sub-themes [20, 21].

The next stage is to develop a prototype based on the issues and needs of the target end users. For example, if a user asks about the inability to obtain weather forecasts effectively, especially in severe weather, and the need to easily map such forecasts with the locations they add; then these can be design and is implemented by adding location associations to the home screens, which can be accessed with a swipe whenever users open their weather app.

The final stage is to evaluate the developed prototype. In particular, we can consider the effectiveness of the prototype by targeting specific populations such as the older. In this case, we need to consider at least two factors: the type of application (i.e. existing vs. new prototypes) and the age group (i.e. young vs. older). Additionally, usability testing can be conducted using several qualitative and quantitative measures, such as completion rates, error counts, error severity ratings, error reasons, completion times, and satisfaction surveys.

### III. UCD PHASE 1: UNDERSTAND USER NEEDS AND LIMITATIONS

**Participants:** In this study, 32 participants were recruited voluntarily, were ( $32 \pm 6.4$ ) years (mean  $\pm$  SD), and ranged in age from 18 to 47 years. Participants were students at the University of Oklahoma and regular users of 11 popular smartphone weather apps running on different operating systems.

**Design:** The participants were guided by skilled moderators and assistant moderators. Several topics related to user perceptions and behaviors of smartphone weather application usability are discussed. These questions are intended to capture general information: (for example, trends in downloading specific weather apps, prioritization of weather apps in critical and non-critical time weather conditions, and positive and negative usability understandings of weather apps), and specific information: (for example, using different location search methods, controlling alert settings, content and display of critical alert messages, and using menu icons and labels).

**Apparatus:** A Nikon D3200 camera was used to record the sessions with the participants. A desktop computer, a projector device, and a large whiteboard surface were used to display the questions to the participants.

**Procedure:** Participants were held in a well maintained controlled environment at the University of Oklahoma. Participants signed a consent form upon arrival and then described the purpose of the study. After that, discussions begin, with each session lasting approximately 90 minutes.

**Data analysis:** Data was collected through video recording sessions. Afterwards, the collected data was transcribed, characterized and classified using qualitative thematic analysis [19]. Specifically, using thematic analysis methods, the data are: 1) transcribed verbatim, 2) encoded using representative words and phrases, and 3) re-fined into common themes.

#### A. Results of Phase 1

Participants presented three main themes: usage efficiency, user cognitive load, and effectiveness. These themes are primarily associated with common features on most popular weather apps (location search, weather forecast, alert messages, alert settings, and radar maps), as detailed below.

1) *Usage efficiency:* Participants were highly concerned with the time and number of steps required to access time-critical information, such as in a weather app. Participants shared several examples of inefficiencies in the design of the location search feature in popular weather apps. For example, in addition to conventional search methods (such as zip and/or city), the widely used weather app Weather Radio has recently adopted a precise location search feature (which enables users to search and save the specific locations) to provide users with accurate weather forecast. The application restricts users from searching for the different locations as required to navigate on the map and identify the desired location (see the example in Fig. 2(a)). Therefore, end users need an effective search method, since this feature needs a great degree of knowledge of the geographic area of the map, widespread visual attention,

and repeated zooming in/out inside the small smartphone screen. Another efficiency requirement has to do with weather forecasting (such as humidity and temperature) feature. Users indicated that a limited number of steps were required because they felt it was accessed more frequently than other features. Several current weather applications involve multiple steps on several screens to access weather forecasts (see examples in Fig. 3(a-d)).

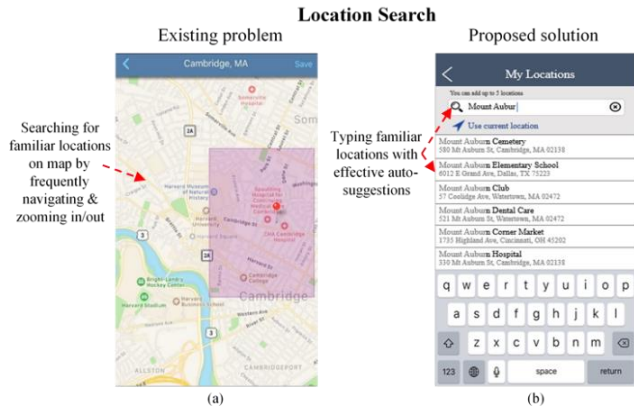


Fig. 2. Screenshot example of a user's main problem with the location search feature and proposed solution. (a) Screenshot of location search in the Weather Radio app, which has efficiency issues where users spend a lot of time navigating and zooming in/out on the map to find familiar locations. (b) Screenshot of a proposed efficiency solution in the UCD app, where the user simply types a familiar location and selects it from a list of auto-suggestions

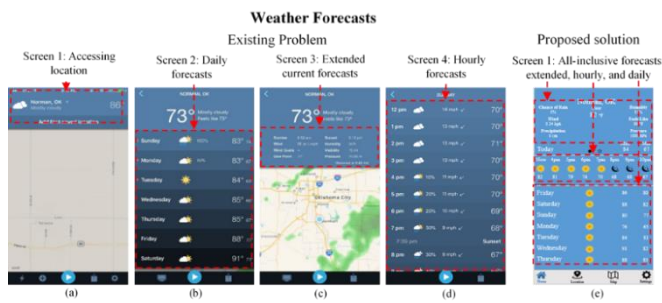


Fig. 3. Screenshot example of a user's main problem with the weather forecasts feature and proposed solution. Screens (a) to (d) are screenshots of weather forecasts in the Weather Radio app with efficiency issues: multi-screen/step forecasts, where the user must access each set of forecasts from a different screen. Screen (e) is a screenshot of the proposed efficiency solution in the UCD app, where users can access predictions for all locations within one screen immediately after opening the app

2) *User cognitive load*: Due to critical situations associated with extreme weather conditions, users also expressed a need for weather apps to simplify their cognitive processes and decision-making strategies. For example, users place high emphasis on the requirement for concise and structured push alert messages throughout severe weather conditions; alerts are created by the weather agency's systems and automatically directed to third parties, including weather apps, where they push them to the end user exactly as they are received. This is because alerts often contain technical data (such as geographic area codes) and cluttered data (see the

example in Fig. 4(a)), which can hinder users' understanding of the alert and rise their cognitive load. Additionally, alerts often contain a huge quantity of evidence; most of this is not related to user-saved locations (e.g. names and information for all alert areas). While retrieving information about distant or isolated locations is critical for some operators in certain usage contexts, such as when traveling, making them a major part of alerts can be detrimental to users. Users also need a radar map feature that is built-in to identify and understand. Participants shared several examples of few popular weather applications that lack intuitive and/or visible instructions on in what way to access maps of the respective saved localities or control their settings (see examples in Fig. 5(a, b)).

3) *Effectiveness*: Owing to the limited smartphones screen size and the critical information displayed, users demand optimized and flexible functionality. For example, in the alert settings feature, the weather app either lets users control of all types of weather alerts and sub-alerts (see example in Fig. 6(a, b)), or does not permit the control of any alerts, but automatically uses active warnings as alerts. For this reason, users have expressed concern about large numbers of weather warnings, most of which are rarely needed and / or insignificant for the average user. Additionally, notifications that push any active alerts without end-user control are considered mandatory interactions. Therefore, users have expressed the need for the ability to control only a few relevant alerts.

#### IV. UCD PHASE 2: DEVELOP PRODUCT PROTOTYPE

A prototype weather app was developed after understanding the specific concerns and needs of users and taking into account widely used smartphone design heuristics, age-related constraints and needs, and common usability guidelines. The prototype application was designed using the InVision application software (<https://www.invisionapp.com/company>) called "EZ Weather".

##### A. Design Proposal Overview

The proposed design solution for the weather app features has been implemented in the EZ Weather app. To address the location search problem, a similar approach to the Google Maps app could greatly improve the efficiency of this feature: enter a familiar location with effective auto-suggestions. With this feature, the user can add an exact location by typing the location name/address (e.g. hospital name/address) in the search bar of the application and selecting it from the auto-suggest list (see example in Fig 2(b)). To reduce the time and number of steps to access weather forecasts, it may be helpful to have all weather forecasts for each saved location in one screen: all-inclusive weather forecast in one screen (see example in Fig 3(e)). For alert messages, a method for filtering the content of the message can significantly reduce the cognitive load on the user by including only relevant and necessary information on the main alert screen that is relevant to the user's saved location; all other information can be accessed from the secondary menu, including distant under-alert area.

In addition, simple language and hierarchically structured information using everyday language can enhance user understanding of messages and responses to alert threats. See Fig. 4(b)), for an example, content that is structured, prioritized, and language simplified. In addition, to facilitate the user's identification and understanding of the radar map feature, it may be helpful to display the list of all saved locations; from which the user can visually select the desired location; then the map of the selected location and its settings are displayed: use the visible and intuitive Map menu (see example in Fig. 5 (c & d)).

To avoid an excessive number of alerts that the user needs to control and interact with, and to give the user the freedom to control the push alert notifications, it may be beneficial to include the most critical and common alerts (on/off) to control. In bad weather, users will receive alert notifications about the time-critical alerts they have turned on (e.g. tornado warnings). Alerts that are not time-critical, such as wind, are not spontaneously sent to the user as notifications; these alerts only appear after clicking the respective symbol on the affected location screen (Fig. 6 (c & d)).

Other key usability and smartphone application guidelines may also need to be considered to improve the overall user experience. These guidelines include labeling all icons with representative text labels, using short descriptive information next to menu options, providing appropriate feedback on user actions, and using appropriate colors with easily distinguishable contrast to visualize and indicate weather conditions (e.g. red screen for locations under a warning alert), provide shortcuts (for example, constant main icons across all screens), text in relatively large fonts, and use a consistent and intuitive layout.

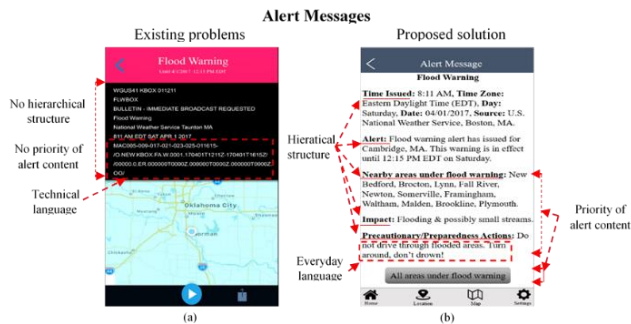


Fig. 4. Screenshot example of a user's main problem with an alert messages feature and proposed solution. Screen (a) is a screenshot of an alert message in the Weather Radio app, showing information about all alert-deficient areas within a scrollable area, uses code and technical language, and without hierarchical information. Screen (b) is a screenshot of the proposed solution in the UCD app, by prioritizing alert content (information about under-alert saved location and nearby areas on main screen and distant under-alert areas accessed from a secondary screen), using everyday language, and with structure hierarchical information

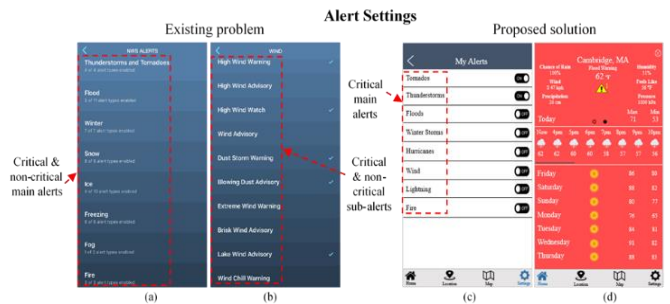


Fig. 6. Screenshot example of a user's main problem with alert settings feature and proposed solution. Screens (a) & (b) are screenshots of alert settings in the Weather Radio app, where users are required to scroll up/down and control the settings of all-weather types' main alerts and sub-alerts (e.g. wind alert has 16 sub-alerts), including critical and non-critical alerts. Screens (c) & (d) are screenshots of a proposed solution of minimalist alert settings in the UCD app, where users need to control the settings of only critical alerts as in (c); non-critical alerts' information can be directly accessed from screen of affected location (e.g. by tapping yellow alert icon as in (d))

## V. UCD PHASE 3: EVALUATE PROSPECTIVE USERS' INTERACTIONS WITH PROTOTYPE

### A. Method of Phase 3

Participants: In this study, eighty (40 young and 40 elder) regular users of the iOS smartphone weather app (first-time users of the tested app) were voluntarily recruited for the experiment. The younger and older participants were  $25.9 \pm 4.8$  years (mean  $\pm$  SD) and  $57.4 \pm 4.3$  years, respectively, and ranged in age from 18 to 35 years and 50 to 66 years, respectively. The participants were assigned randomly to perform tasks in two test apps (Weather Radio and EZ Weather). Both applications were used by 40 users (20 young and 20 elder). Recruitment was based on personal communications, mass emails from universities and flyers hanging on doors of various public buildings. Participants signed a consent form upon arrival and then described the purpose of the study. The experimental protocol was approved by the ethical committee of the department of Human Research Participant Protection (HRPP), University of Oklahoma, Norman campus with IRB: 6681/2021.

Apparatus: For the present study, both the applications i.e., Weather Radio (version 3.0.5) and EZ Weather (version 1.0.0) are installed on iOS Smartphone. Using high fidelity simulations (InVision), a powerful interaction design system, we displayed recorded alert messages on the Weather radio at any time during the experiment. To record user interactions, in particular to calculate the time to complete a given task, to

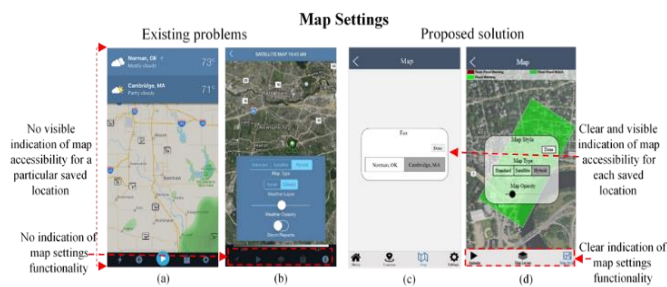


Fig. 5. Screenshot example of a user's main problem with radar/map feature and proposed solutions. Screens (a) & (b) are screenshots of radar/map in Weather Radio app that includes visibility and recognition issues, as neither indication is available on how to access map of a desired saved location as in (a) nor the functionality of icons is easily recognized as in (b). Screens (c) & (d) are screenshots of proposed solutions in the UCD app, where (c) shows all saved locations for a user to select and access map of a desired location and (d) provides icon labels to ease functionality recognition

count and classify errors the Nikon L340 camera was used. Demographics, post-task and post-test surveys were printed on paper.

**Scenario & Tasks:** Participants were given a scenario in which their grandmother was admitted to Mount Auburn Hospital in Cambridge, MA, and noticed that a flood warning had been announced in the Cambridge area. To understand grandmother's exact location alert risk level and access all relevant information, participants can search for and add to grandmother's exact location, access and identify relevant weather forecasts and alert messages. Further, participants need to view alerts on map in map settings and adjust alert settings to receive relevant alert notifications. Fig. 7 shows the successful completion of the scenarios and tasks.

**Procedure:** First, participants were explained the purpose of the survey, signed a written consent form, and then completed the demographic survey. None of the participants received training and were not given the opportunity to practice themselves using the test app; the participants were given scenario and task guidelines and were asked to start the experiment after they were told they were ready. Weather Radio is a running application, so participants won't see any alert messages unless during an active alert, so Weather Radio's alert message assignment was presented to participants through a smartphone interface designed by InVision. The alert message displayed to participants is a flood warning posted to Weather app users in Cambridge, MA on April 1, 2017. After finishing each task, participants were asked to complete a post-task survey to provide their instant feelings about each task/feature. Finally, participants were asked to complete a post-test survey to evaluate their satisfaction with the test app.

**Experimental Design & Variables:** Full factorial design (2 levels of app \* 2 levels of age group) were used in the present study as independent variables. The dependent variables were effectiveness, efficiency, and user satisfaction. The control variable was user experience; all users in both age groups must have at least six months of experience with iOS smartphones and the Weather app (excluding the app tested). Error metrics are the reasons for errors made (usability issues), how often they occur, the proportion of users who make mistakes, and the severity level of the cause of the error. The error severity ratings used in this study are based on a rating scale proposed by Nielsen [23] (Table I).

Finally, Pearson correlation test (r) was also performed to define the relationship between the usability measures used in this study.

The post-task survey was done by using a Single Ease Question (SEQ) with 7-point Likert rating scale [22]. This question was used because it was found to be as effective as other complex measures of task difficulty, namely the Usability Magnitude Estimate (UME) and Subjective Mental Effort (SMEQ) questionnaires [23]. The post-test survey was done by using User Interface Satisfaction (QUIS) questionnaire [24].

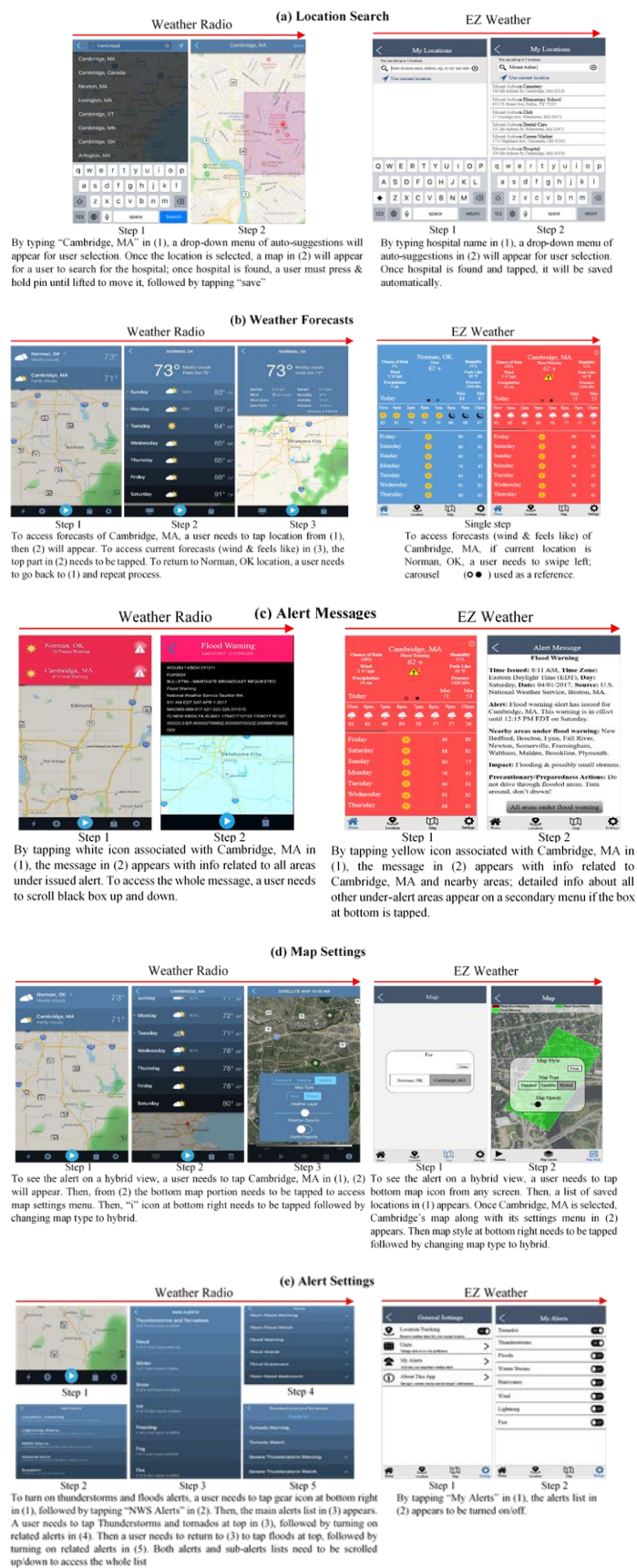


Fig. 7. (a-e). Process of performing tasks on Weather Radio vs. EZ Weather

Data Analysis: All data's were analyzed using SPSS version 23. A two-way analysis of variance (ANOVA) between subjects was performed to determine the effect of both application used and age group task completion time for all tasks. The Mann-Whitney U test was used to examine the size of the differences between the app and age group variables used, both in terms of the number of errors and the metrics of the post-task satisfaction survey. For post-test satisfaction surveys, mean and standard deviation errors were calculated.

TABLE I. NIELSEN'S SEVERITY RATING SCALE OF THE USABILITY PROBLEMS

Rating	Nature of usability problems
0	I don't agree that this is a usability problem at all
1	Cosmetic problem only: need not be fixed unless extra time is available on project
2	Minor usability problem: fixing this should be given low priority
3	Major usability problem: important to fix, so should be given high priority
4	Usability catastrophe: imperative to fix this before product can be released

B. Results of Phase 3

1) Effectiveness

a) Completion rate: Fig. 8 shows that all users in both groups were able to complete a given task in EZ Weather successfully. However, several users were unable to complete all three tasks on Weather Radio, and older users had a higher failure rate on the location search task.

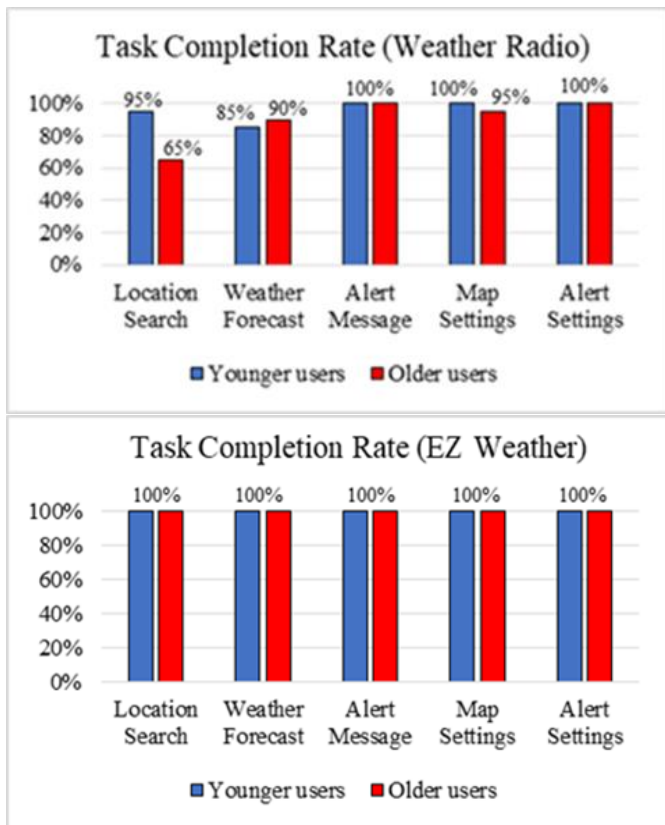


Fig. 8. Proportions of successful task completion rate for both app

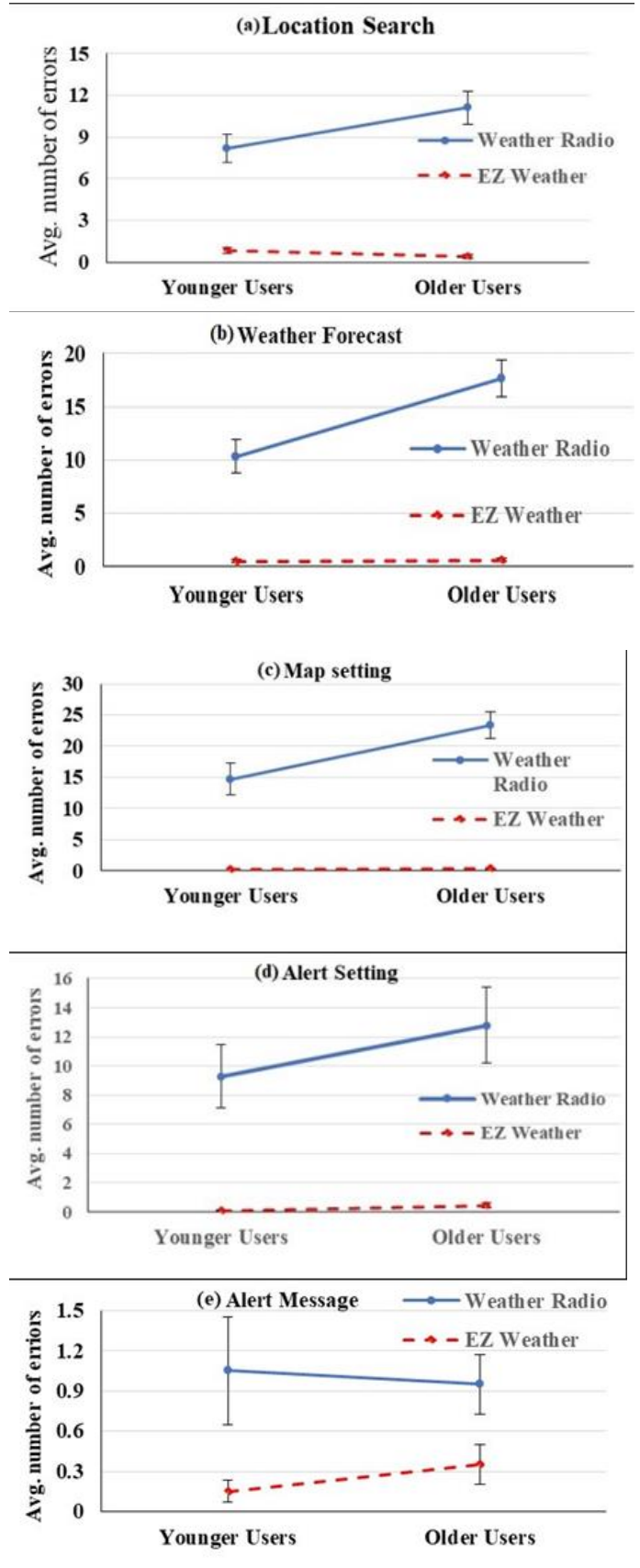


Fig. 9. (a-e). Average number of errors using both apps

TABLE II. MAN-WHITNEY TEST SUMMARY FOR NUMBER OF ERRORS

Task	Source	Z-score	U-test	P-value
Location Search	Age group	-1.22	677.50	0.223
	App used	-7.86	9	<0.001
Weather Forecast	Age group	-2.14	582	0.033
	App used	-6.64	123	<0.001
Alert Message	Age group	-.56	743	0.575
	App used	-7.68	20	<0.001
Map Settings	Age group	-2.42	655.5	0.015
	App used	-7.51	34.5	<0.001
Alert Settings	Age group	-.91	710.5	0.364
	App used	-8.01	10	<0.001

TABLE III. CAUSES OF ERRORS, FREQUENCY OF ISSUE, (PROPORTIONS OF USERS WHO MADE ERRORS), AND AVERAGE SEVERITY RATINGS ON WEATHER RADIO

Feature	Cause of errors (Usability problem)	Frequency of issues		Average Severity Rating
		Young Users	Older Users	
Location Search	Users are having trouble finding locations and moving pins on the map.	178	253	3.5
Weather Forecast	Users couldn't easily find the weather forecast. The area that causes the prediction, if clicked, appears to be unclickable.	231	366	3.5
Alert Message	Users couldn't easily access the necessary information of the time-critical weather alert message because of cluttered & unstructured information and poor use of language.	21	17	4
Map Settings	Users struggle to start the task due to counterintuitive steps and an invisible map settings menu.	166	287	3
Map Settings	Users are unaware of the functionality of the map settings icons because these icons are neither marked nor standardized across mobile applications.	126	194	2.5
Alert Settings	Users do not understand the function of home screen icons because these icons are neither marked nor standardized in smartphone applications. Also, the large number of alerts and sub-alerts seems to confuse users about the options required for the task.	186	256	3.5

b) *Errors*: Fig. 9(a-e) shows that errors on EZ Weather are significantly less than those on Weather Radio for all tasks, for both young and older users. The results showed that the older users had made more mistakes on both the apps compared to younger participants on all tasks except Weather Radio's alert message task and EZ Weather's location search

task. The Mann-Whitney U test showed a significant difference ( $p < 0.001$ ) in the number of errors across all tasks between the two apps (Table II). The results showed no significant error differences for all tasks between the both groups, except for the weather forecast and map setting tasks. For other error-related metrics, almost all users in both groups made errors, and with different frequencies, due to the usability issues with every task in Weather Radio except the Alert Messages task (Table III). Nearly half of the participants made mistakes due to alert message usability issues. In contrast, fewer users made mistakes in EZ Weather's tasks, with a much lower frequency and average severity rating, compared to Weather Radio. Errors on EZ Weather are mainly caused by swipe actions (such as spelling mistakes and accidental clicks on adjacent function icons) (Table IV).

TABLE IV. CAUSES OF ERRORS, FREQUENCY OF ISSUES, (PROPORTIONS OF USERS WHO MADE ERRORS), AND AVERAGE SEVERITY RATINGS ON EZ WEATHER

Feature	Cause of errors (Usability problem)	Frequency of issue		Average Severity Rating
		Young Users	Older Users	
Location Search	Users made typing errors when typing location name.	17	9	0
Weather Forecast	Users couldn't easily figure out that accessing weather forecasts of different locations was through swiping the screen right or left.	10	12	1.5
Alert Message	Users didn't expect the alert message icon to be clickable and/or required to access the message, when clicked.	3	7	1
Map Settings	Users mistakenly tapped adjacent icons of unrelated functions.	4	6	0
Alert Settings	Users mistakenly tapped adjacent icons of unrelated functions.	2	9	0

2) *Efficiency (task time)*: Fig. 10(a-e) shows that, on an average, the younger and older participants take fewer time to accomplish each task of EZ Weather compared to Weather Radio. The two age groups had relatively taken same times on all tasks for both apps. Therefore, two-way ANOVA (Table V) shows that on both weather apps, there were no significant time differences between young and elder users for all tasks except the map setting task. In the weather app tested, however, the time to complete all tasks was significantly different ( $p < 0.001$ ) between the both groups. In addition, there was no significant interaction between the both group and the app used for all tasks.

3) *Post-task satisfaction ratings*: Fig. 11(a-e) shows that both young and older participants found the EZ Weather task easier than the Weather Radio task. Younger users rated the ease of use for both apps higher compared to older participants, though the difference was not huge. In addition, the Mann-Whitney U test showed a statistically significant difference ( $p < 0.001$ ) in task satisfaction between the two apps



for all tasks (Table VI). It also showed that young and older users were similarly satisfied with all tasks.

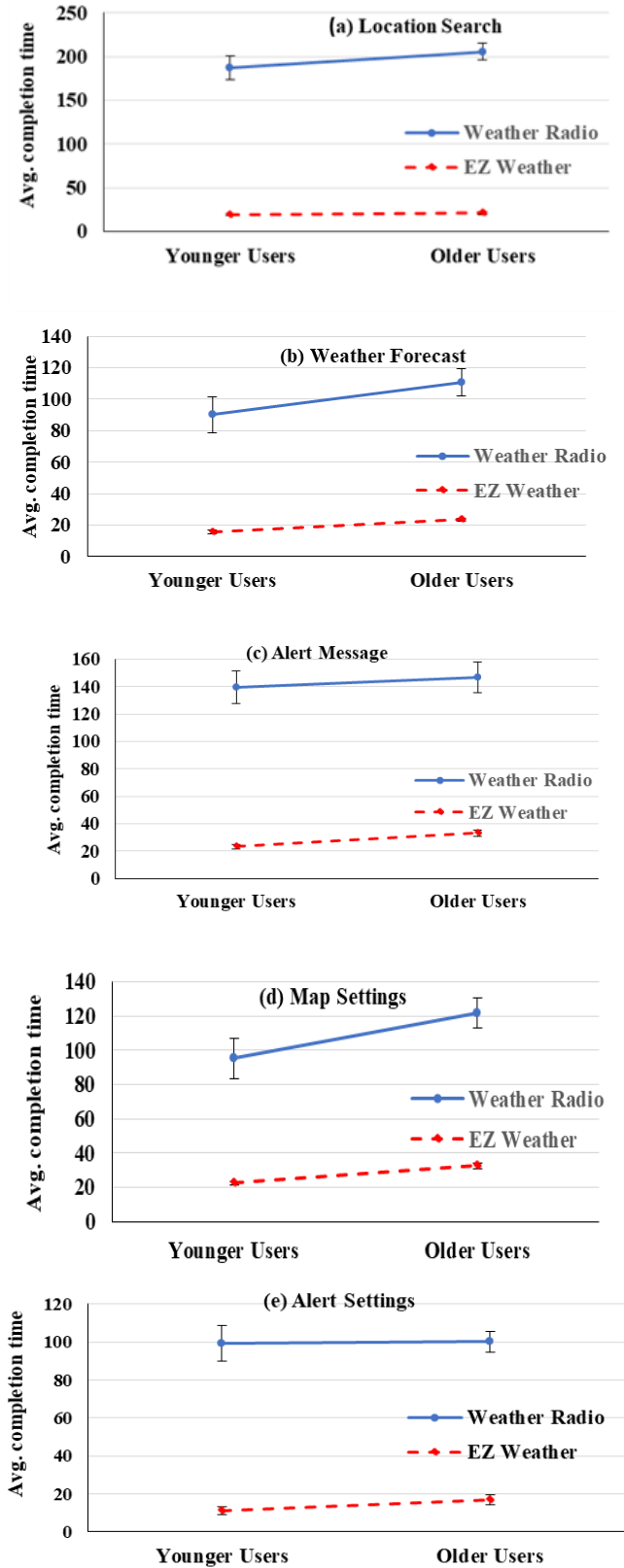


Fig. 10. (a-e). Average time spent (in seconds) in completion of tasks

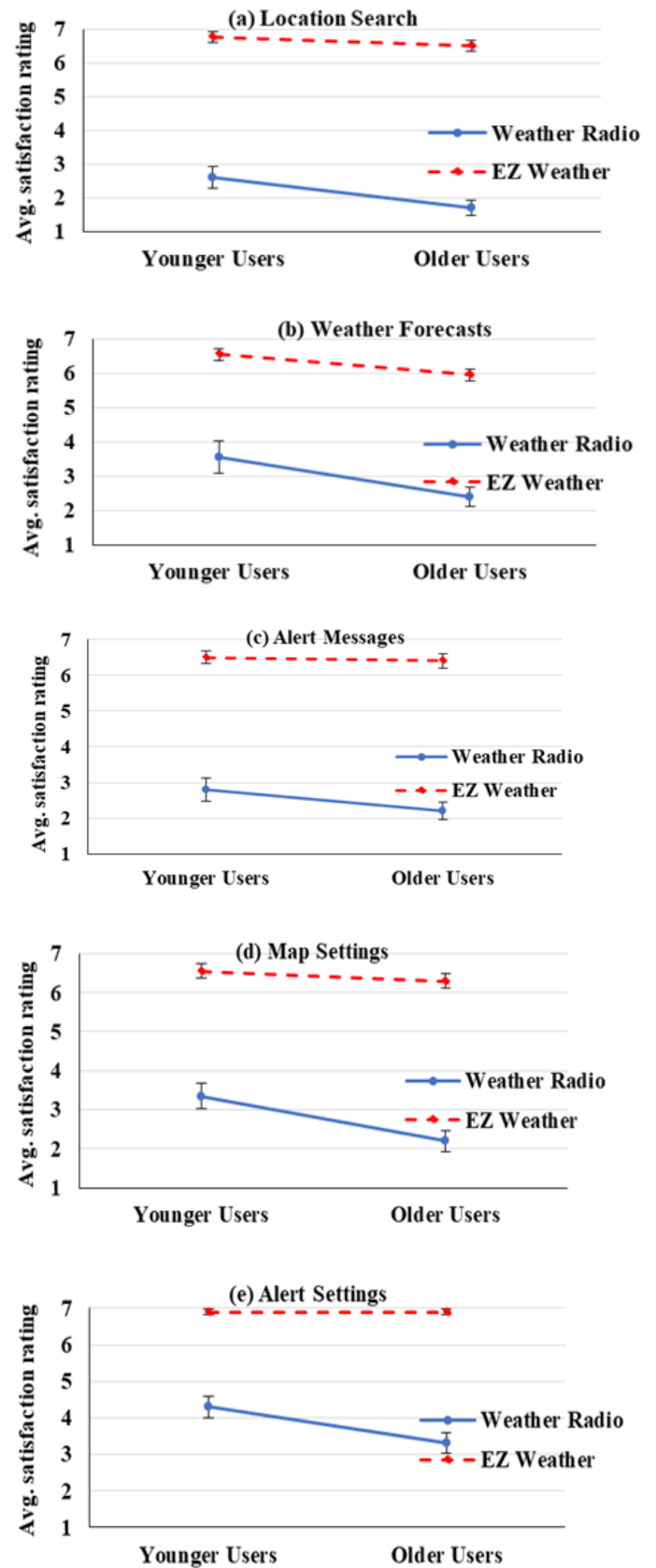


Fig. 11. (a-e). Average post-task satisfaction ratings of SEQ survey

TABLE V. TWO-WAY ANOVA SUMMARY FOR TASK TIME

Task	Source	F	P-value
Location Search	Age group	1.401	0.240
	App Used	445.554	<0.001
	User * App	0.949	0.333
Weather Forecasts	Age group	3.916	0.051
	App Used	127.384	<0.001
	User * App	0.805	0.373
Alert Messages	Age group	1.010	0.318
	App Used	189.608	<0.001
	User * App	0.023	0.879
Map Settings	Age group	6.012	0.017
	App Used	116.856	<0.001
	User * App	1.205	0.276
Alert Settings	Age group	0.332	0.566
	App Used	242.835	<0.001
	User * App	0.194	0.661

TABLE VI. MAN-WHITNEY TEST SUMMARY FOR POST-TASK SATISFACTION RATINGS

Task	Source	Z-score	U-test	P-value
Location Search	Age group	-0.298	769.5	0.766
	App used	-7.76	5	<0.001
Weather Forecast	Age group	-1.32	665	0.187
	App used	-7.58	25	<0.001
Alert Message	Age group	-0.842	727.5	0.401
	App used	-2.68	569	0.007
Map Settings	Age group	-1.33	667	0.184
	App used	-7.97	1.5	<0.001
Alert Settings	Age group	-1.32	668	0.185
	App used	-7.09	93.5	<0.001

4) *Post-test satisfaction ratings:* Overall, the results in Fig. 12 and 13 was based on the QUIS questions [24], where the same questions were asked to both groups of participants under same parameters, suggesting that younger and older participants feel comparatively similar to the interface de-sign specifications, and younger users are more satisfied. In addition, all participants were very pleased with EZ Weather, while most users of Weather Radio were disappointed. Users report that the screen design of EZ Weather is very good, the text and images are very clear, the fonts are very clear, the amount of information presented is sufficient, and the arrangement is reasonable. However, participants were

appeared to be disappointed with the amount of information displayed on Weather Radio; they observed that there was too much information, particularly about time-critical alerts.

Users have found consistent terminology and messaging throughout EZ Weather as well as the location of on-screen instructions. Participants reported that performing any action produced predictable results in EZ Weather, with a very acceptable delay between actions. However, many participants noted that they were not sure what to expect when performing multiple actions in Weather Radio. Compared to Weather Radio, users report that using EZ Weather is very easy to pick up, learn, and efficient. Additionally, users seem to be very pleased with the number and order of steps required to accomplish each EZ Weather task. They also liked the feedback they received as they completed each task. Navigating features and remembering names and commands was also easy for participants in both groups. When it comes to the EZ Weather feature, although many users made no mistakes during the experiment, they reported that errors and typos were easy to correct. Users also like their ability to perform or undo actions using shortcuts. In addition, they say that users of any experience level can easily and consistently complete their tasks. The choices of color for both apps seem natural enough for most users, with EZ Weather having a higher user satisfaction rate.

5) *Correlations among usability measures:* Tables VII and VIII show the Spearman rho correlation ( $\rho$ ) results among all the indicators in this study. Since all users successfully performed a given task, the correlation between completion rates for the two age groups and other metrics on EZ Weather was not performed. The results showed a strong positive correlation between task time and errors, and between post-task and post-test satisfaction, for both age groups for both applications. However, the time spent on each task and errors were negatively correlated with post-task and post-test satisfaction scores.

TABLE VII. CORRELATION MATRIX FOR BOTH YOUNGER AND OLDER USERS ON WEATHER RADIO, IN TERMS OF P & (P-VALUES).

	Younger User			
	Time	Errors	Task-Sat	Test-Sat
<b>Completion rate</b>	-0.615 (0.002)	-0.682 (0.003)	0.766 (<0.001)	0.870 (<0.001)
<b>Time</b>		0.885 (<0.001)	-0.921 (<0.001)	-0.934 (<0.001)
<b>Errors</b>			-0.858 (<0.001)	-0.898 (<0.001)
<b>Task-Sat</b>				0.917 (<0.001)
Older User				
<b>Completion rate</b>	-0.651 (0.003)	-0.792 (<0.001)	0.784 (<0.001)	0.921 (<0.001)
<b>Time</b>		0.907 (<0.001)	-0.905 (<0.001)	-0.908 (<0.001)
<b>Errors</b>			-0.843 (<0.001)	-0.929 (<0.001)
<b>Task-Sat</b>				0.911 (<0.001)

TABLE VIII. CORRELATION MATRIX FOR YOUNGER AND OLDER USERS ON EZ WEATHER, IN TERMS OF P & (P-VALUES).

		Younger User		
		Errors	Task-Sat	Test-Sat
Time		0.798	-0.883	-0.844
		(<0.001)	(0.002)	(<0.001)
Errors			-0.704	-0.711
			(<0.001)	(<0.001)
Task-Sat				0.841
				(<0.001)
		Older User		
		Errors	Task-Sat	Test-Sat
Time		0.781	-0.820	-0.851
		(<0.001)	(<0.001)	(<0.001)
Errors			-0.798	-0.831
			(<0.001)	(<0.001)
Task-Sat				0.892
				(<0.001)

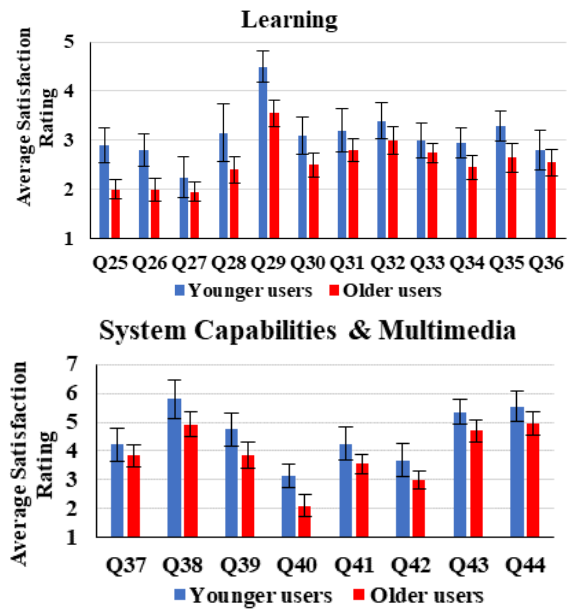
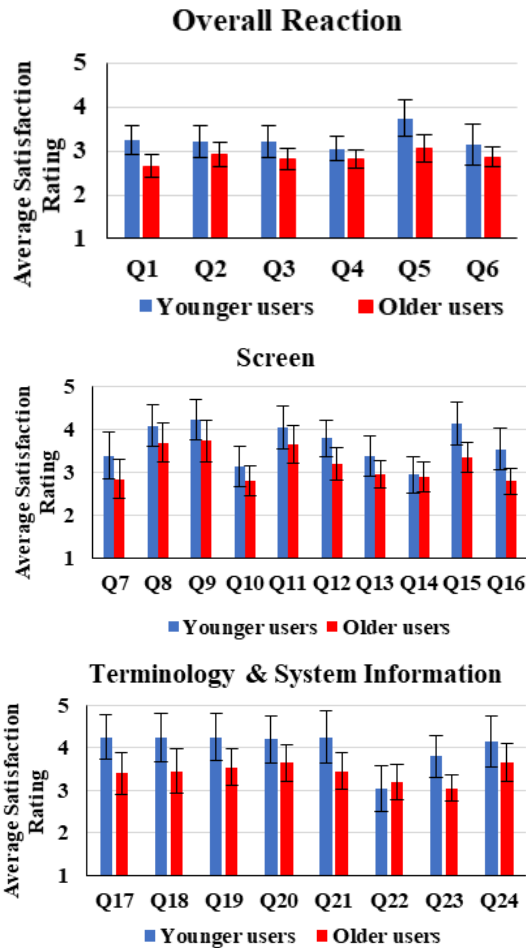
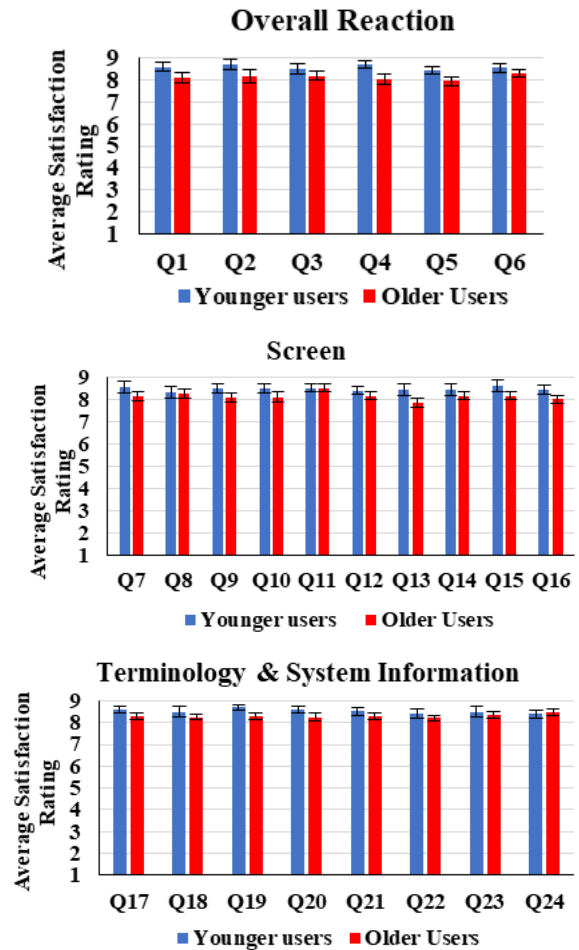


Fig. 12. Average post-test satisfaction rating of QUIS survey (Weather Radio)



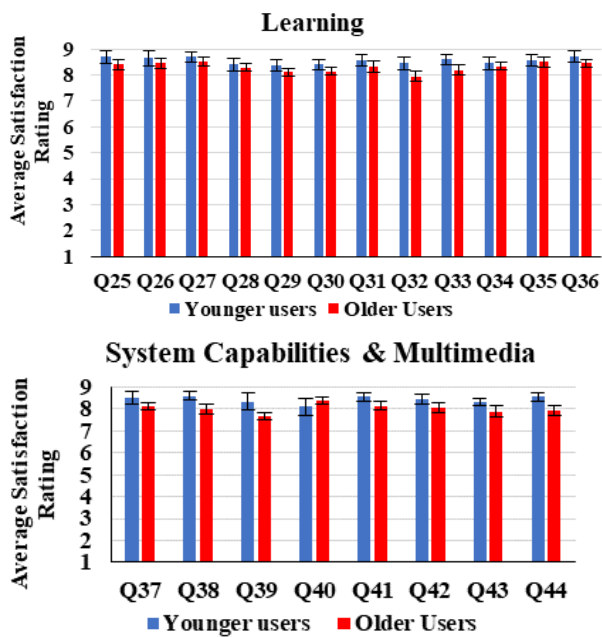


Fig. 13. Average post-test satisfaction rating of QUIS survey (EZ Weather)

## VI. DISCUSSION

Overall, this study establish that the UCD weather app (EZ Weather) was considerably more useful compared to the other popular weather app (Weather Radio) for all the evaluation metrics used. Also, younger and older participants seem to have comparable results on all tasks for both the weather application, with younger users experiencing marginally greater performance and satisfaction. These results show that regard-less of age, an application's interface design significantly affects end-user performance and perceptions (positive or negative) of application usability. The results further demonstrate that adopting a UCD approach to applications that contain time-critical data, such as weather applications, will result in extremely interactive and usable systems. Results for all metrics had showed that prioritizing and structuring critical evidence and using everyday terms throughout the EZ Weather interface greatly helps participants easily interacting with inherent features and finds them beneficial. For example, all young and older participants were able to complete alert message tasks in EZ Weather with far greater efficacy and task satisfaction compared to Weather Radio features.

Additionally, only a handful of users had errors in some of the usability issues category in EZ Weather's alert messages task, while nearly half of participants of each group had errors in this task for Weather Radio due to usability issues in the catastrophe category. In addition, the alert message task further demonstrate that the push alerts will greatly help end participants to correctly perceive alert threats and respond effectively to alert threats, particularly in the forthcoming weather conditions.

The greater usability of the EZ Weather is also due in part to consideration of efficiency principles during the design phase. This is supported by the very limited time and steps required to access any feature on EZ Weather. A good example

is the weather forecast feature that includes all weather prediction information for respectively saved location within same screen (all-inclusive weather forecast); with just one swipe right or left, the user can access other saved locations' weather forecast. Although some users couldn't figure out the swipe functionality, from the first test, all success-fully completed the weather forecast task with significantly greater efficacy and task satisfaction, with fewer and less errors, than Weather Radio multiple-screen weather forecasts feature. Another example of an efficient feature of EZ Weather is the location search feature. With efficient auto-suggestions to type in familiar locations, both groups of users can complete the corresponding task in seconds. By contrast, navigating maps and pinpointing locations within the limited smartphone screen size is not only inefficient, but also ineffective, especially for older users, as demonstrated by Weather Radio's high failure rate and completion time for location searches.

Users' greater performance and satisfaction with EZ Weather than Weather Radio may also be related to EZ Weather's simple design. For example, limiting user-controlled alerts to the most common and critical alarms on EZ Weather (using a minimalist alarm setup) enables users to effectively and efficiently perform alarm set-up tasks with great satisfaction. Moreover, EZ Weather also permits participants to use non-critical alerts (via an illustrative icon displayed on the exaggerated location screen) during active alerts. The significant drop in performance and satisfaction with Weather Radio's alerts setting tasks was primarily because of the huge number of alerts and sub- alerts that users need to navigate and control (e.g., 16 sub- alerts for wind alerts alone). Two other important principles considered in the design of the EZ Weather interface are its usability and ease of recognition. For example, while the map setting task on both apps requires interaction with similar step and function icons. EZ Weather's greater performance and satisfaction on this task is supposed to be recognized to the visualization and intuitive menus (using the visualization and intuitive map menu), and appropriately labeled icons. These factors and results show that the actions have been performed with expectable results in a logical sequence. However, invisible elements in smartphone apps (such as the map settings menu bar) and icons that are neither marked nor standardized may be one of the main reasons for the dramatic drop in performance and satisfaction with Weather Radio tasks.

The usability and design principles monitored in the design of EZ Weather that may have ultimately contributed to great results include: feedback on system status (e.g. confirmation messages to perform activities), uniformity of application elements (e.g. settings menu), shortcuts usability to expedite interaction and correct errors, provide short descriptive information to help users understand the functionality of the corresponding feature, and use large text fonts along with high contrast and indicative colors to address age-related limitations for older users. The post-test satisfaction survey results were consistent with those of the task-based metrics. The two apps differed significantly in satisfaction with all interface standards across the two age groups, suggesting that applying usability guidelines to interface design not only results in high performance but also satisfaction for users. Also, on most

QUIS survey items, the two age groups are highly similar in satisfaction, meaning that age differences do not significantly affect user satisfaction. The strong correlations between all used metrics complement the results of previous study [25], which suggest that usability metrics are interdependent aspects that inform the usability of an interface. However, Frøkjær et al. [26] showed the dependence of usability metrics depends on whether the tested interface contains highly complex functionality. If the domain of interest includes complex features, a weak correlation is expected; if not, a strong correlation is most likely.

## VII. CONCLUSION

This study validated the usability of the UCD weather app (EZ Weather) by comparing it to a popular app (Weather Radio). In addition, this study also considered young and older users to determine whether the usability of the test app would be highly influenced by age differences. Results for all metrics showed that users' performance and satisfaction on EZ Weather improved significantly regardless of age, demonstrating the importance of considering UCD methods, usability heuristics, and smartphone design rules in interface development.

## ACKNOWLEDGMENTS

The authors would like to thank all team members and support staff who were directly or indirectly involved during the development and testing of this application.

## REFERENCES

- [1] P. Lawton. "Aging and performance of home tasks". *Human factors*. 1990, 32(5): 527-536.
- [2] J. Sweller. "Cognitive load during problem solving: effects on learning". *Cognitive Science*. 1988, 12(2): 257-285.
- [3] S. Czaja. "Factors predicting the use of technology: Findings from the center for research and education on aging and technology enhancement". *Psychology and Aging*. 2006, 21(2): 333-352.
- [4] M. Khawaji. "Towards an enhanced user experience with smart phone weather alert applications: Usability and User-Centered Design approaches", 2018. Retrieved from <https://hdl.handle.net/11244/302100>.
- [5] T. Levdikova. "Designing apps for elderly smartphone users", 2019. Retrieved from <https://clutch.co/app-development/resources/designing-apps-for-elderly-smartphone-users>
- [6] F. García-Peñalvo, M. Conde and O. Matellán. "Mobile apps for older users—The development of a mobile apps repository for older people", in *Proceedings of the International Conference on Learning and Collaboration Technologies 1st Annual Meeting*, Heraklion, Greece, 2014.
- [7] A. Morey, H. Barg-Walkow and A. Rogers. "Managing heart failure on the go: Usability issues with mHealth apps for older adults", in *Proceedings of the Human Factors and Ergonomics Society 61st Annual Meeting*, Los Angeles, 2017.
- [8] L. Smith, S. Chaparro. "Smartphone text input method performance, usability, and preference with younger and older adults". *Human Factors: The Journal of Human factors and Ergonomics Society*. 2015, 57(6): 1015-1028.
- [9] A. Khamaj, Z. Kang. "Usability evaluation of mobile weather hazard alert applications". *Industrial and Systems Engineering Review*. 2018, 6(1): 21-40.
- [10] J. Drogalis, E. Keyes and L. Dhyani. "The Weather Channel application usability test", 2015. Retrieved from <https://elkeyes.com/wp-content/uploads/2015/05/WeatherChannelUTResults.pdf>
- [11] International Organization for Standardization, *Human-centered design for interactive systems (ISO 9241-210:2010)*, 2010.
- [12] D. Norman. *The design of everyday things: Revised and expanded edition*. New York, NY: Basic Books, 2013.
- [13] Y. Mao, K. Vredenburg, W. Smith and T. Carey. "The state of user-centered design practice". *Communications of the ACM*. 2005, 48(3): 105-109.
- [14] R. Schnall, M. Rojas, S. Bakken, W. Brown, M. Carry and J. Travers. "A user-centered model for designing consumer mobile health (mHealth) applications (apps)". *Journal of biomedical informatics*. 2016, 60(2): 243-251.
- [15] S. Jung, H. Kim, H. Lee. "Influences of perceived product innovation upon usage behavior for MMORPG, Product capability, technology capability, and user-centered design". *Journal of Business Research*. 2014, 67(10): 2171-2178.
- [16] J. Nielsen. "10 usability heuristics for user interface design", 2020. Retrieved from <https://www.nngroup.com/articles/ten-usability-heuristics/>
- [17] Y. Rogers, H. Sharp, and J. Preece. *Interaction design: beyond human-computer interaction*, Hoboken, NJ: John Wiley & Sons, 2011.
- [18] J. Gove. "Principles of mobile app design: Engage users and drive conversions", 2022. Retrieved from <https://www.thinkwithgoogle.com/future-of-marketing/creativity/principles-of-mobile-app-design-engage-users-and-drive-conversions/>
- [19] E. Boyatzis. *Transforming qualitative information: Thematic analysis and code development*, Thousand Oaks, CA: Sage Publications, 1998.
- [20] V. Braun, V. Clarke. "Using thematic analysis in psychology". *Qualitative Research in Psychology*. 2006, 3(2): 77-101.
- [21] J. Nielsen. "Severity ratings for usability problems", 1994. Retrieved from <https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/>
- [22] S. Yekta-Michael, C. Färber and A. Heinzel. "Evaluation of new endodontic tooth models in clinical education from the perspective of students and demonstrators". *BMC Medical Education*. 2021, 21(1): 447.
- [23] J. Sauro, S. Dumas. "Comparison of three one-question, post-task usability questionnaires", in *Proceedings of the Special Interest Group on Computer-Human Interaction conference on human factors in computing systems 27th Annual Meeting*, New York, 2009.
- [24] P. Chin, A. Diehl and L. Norman. "Development of an instrument measuring user satisfaction of the human-computer interface", in *Proceedings of the Special Interest Group on Computer-Human Interaction conference on human factors in computing system*, New York, 1988.
- [25] S. Joo. "How are usability elements—efficiency, effectiveness, and satisfaction—correlated with each other in the context of digital libraries", in *Proceedings of the American Society for Information Science and Technology 73rd Annual Meeting*, Hoboken, 2010.
- [26] E. Frøkjær, M. Hertzum and K. Hornbaek. "Measuring usability: are effectiveness, efficiency, and satisfaction really correlated", in *Proceedings of the Special Interest Group on Computer-Human Interaction conference on human factors in computing system*, 18th Annual Meeting, New York, 2000.

# Interventional Teleoperation Protocol that Considers Stair Climbing or Descending of Crawler Robots in Low Bit-rate Communication

Tsubasa Sakaki<sup>1</sup>, Kei Sawai<sup>2</sup>

Graduate School of Engineering, Toyama Prefectural University  
5180 Kurokawa, Imizu City, Toyama 939-0398, Japan

**Abstract**—In teleoperation of a crawler robot in a disaster-stricken enclosed space, distress of the crawler robot due to communication breakdown is a problem. We present a robot teleoperation system using LoRaWAN as a subcommunication infrastructure to solve this problem. In this system, the crawler robot is operated by teleoperation using a subcommunication infrastructure in a place where a wireless local area network (LAN) communication is possible. In this study, we assume an environment in which the crawler robot must ascend and descend stairs to evacuate to a place where wireless LAN communication is possible. In addition, the disaster-stricken environment is considered as an environment where obstacles are expected to suddenly occur, and the crawler robot has difficulty avoiding obstacles on the stairs. In this paper, we propose a teleoperation communication protocol that considers the risk of sudden appearance of obstacles and confirm its effectiveness using evaluation experiments in a real environment.

**Keywords**—LoRaWAN; teleoperation; crawler robot; disaster-reduction activity; teleoperation protocol

## I. INTRODUCTION

When a large-scale disaster occurs, information gathering is required to assess the damage in the affected area [1]. The methods of information gathering in affected areas include aerial photography using drones, monitoring of existing infrastructures, and onsite activities by rescue workers [2-4]. However, collecting information from the sky in an enclosed space is difficult. Existing infrastructure may become inaccessible due to malfunction or lack of power supply. When rescue workers search a disaster-stricken enclosed space, human casualties due to collapsed buildings and debris may be expected. Therefore, information gathering using mobile robots has been considered in enclosed spaces after disasters [5-7].

Two types of communication methods that use mobile robots are available: wired and wireless. The wired system offers the advantages of stable power supply to the mobile robot via cable and maintains communication quality between the operator and mobile robot [8]. However, teleoperation of the mobile robot using a wired system suffers from problems such as cable tangling and snagging as well as cable handling at the time of return. On the other hand, the wireless method does not reduce the operability due to cables, and mobile robots can maintain high operability. However, as the mobile robot moves, the distance from the base station increases, and the electric-field-strength decreases. As a result, the mobile

robot faces the risk of going into locations where radio waves cannot reach and communication is interrupted. When teleoperating a mobile robot, considering the merits and demerits of wired and wireless communication methods and deciding on a communication method that suits a disaster-stricken environment are important [9-10]. In the present study, we present a wireless teleoperation method for a mobile robot in environments where the mobile robot has difficulty exploring using wired communication.

Related research on information gathering using mobile robots in enclosed spaces includes the robot wireless sensor networks (RWSNs) (Fig. 1) [11-12]. RWSN is a method of extending the communication range between the operator and mobile robot by installing a sensor node (SN), which consists of wireless repeaters, on the mobile robot and deploying them along its movement path. Whereas teleoperation using RWSN enables a wider range of information gathering, the mobile robot may be isolated due to communication breakdown caused by SN failure, battery failure, and radio interference. Distress of the mobile robot can not only interrupt information gathering but also cause secondary disasters because a mobile robot in distress becomes an obstacle or may ignite from its onboard battery. Therefore, in the present study, we present a teleoperation method using a subcommunication infrastructure as a distress-prevention method when the main communication infrastructure [wireless local area network (LAN)] is disconnected.

In the distress-prevention method that uses a subcommunication infrastructure, the operator PC, SN, and mobile robot are equipped with communication devices for the subcommunication infrastructure, and the network is constructed by letting them communicate in a multistage relay. The operator then uses the network to teleoperate the mobile robot and evacuates it to the place where wireless LAN communication is possible.

Subcommunication infrastructure is described as a network that is built based on a wireless-communication standard different from that of wireless LAN. Further, it must be able to continue communication even in an environment where wireless LAN is disconnected. Therefore, when SN becomes unavailable due to SN or battery failure, capability to jump over the failed SN is necessary to ensure continuous communication between the operator and mobile robot. Thus, the subcommunication infrastructure should be a wireless

communication standard that is superior in terms of communication range compared with wireless LAN. In addition, the subcommunication infrastructure must be a communication standard in a frequency band that is different from that of wireless LAN to prevent communication disconnection due to radio interference with wireless LAN. Therefore, in the present study, we realize the proposed method using LoRaWAN, which satisfies the abovementioned conditions as a subcommunication infrastructure (Fig. 2).

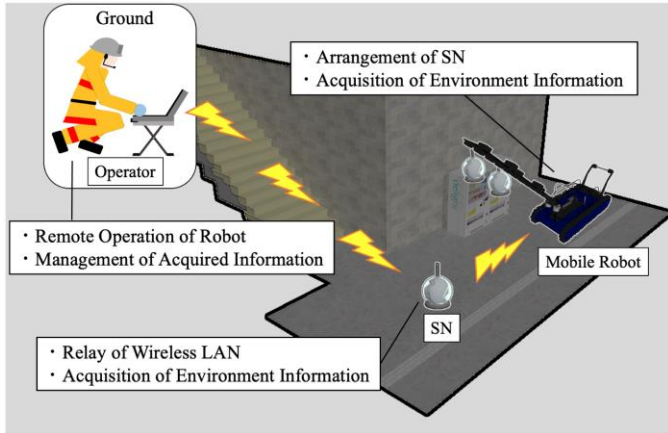


Fig. 1. Information-gathering system using a mobile robot (RWSN)

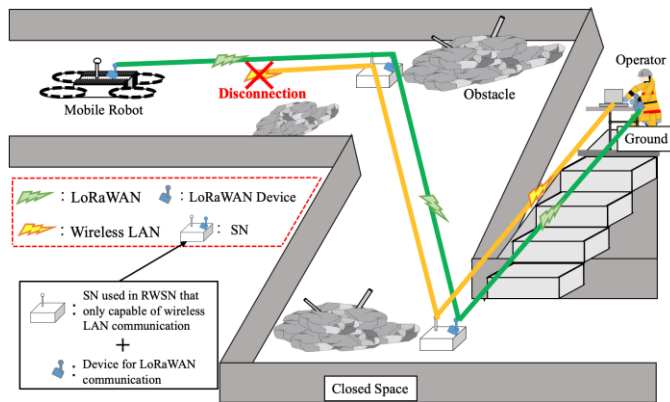


Fig. 2. Distress-prevention methods of mobile robots by teleoperation using LoRaWAN

In the disaster-stricken enclosed space assumed in this study, sudden appearance of obstacles such as collapsed debris is expected. Therefore, when an obstacle is found during the evacuation to the place where wireless LAN communication is available, the mobile robot is required to bypass or climb over the obstacles. The slope of the staircase changes when an obstacle appears in the stair environment. Therefore, if the mobile robot runs on the stairs, it faces the risk of falling down due to insufficient hill-climbing ability. In addition, if the mobile robot climbs over an unstable obstacle, a risk exists in which the robot may fall over or the obstacle collapses. Therefore, shortening the teleoperation time and completing the stair climbing-process before sudden obstacles appear during the stair climbing are necessary by teleoperation using LoRaWAN.

In this paper, we propose an interventional teleoperation protocol. The proposed protocol eliminates the factors that

increase the teleoperation time in existing teleoperation protocols. In the experiment, we performed teleoperation using the proposed and existing methods in a real environment and confirmed that the proposed method could reduce the teleoperation time and demonstrated its effectiveness.

## II. LORAWAN OVERVIEW

LoRaWAN is one of the communication standards categorized as low-power wide-area standard, which is a generic term for wireless-communication technologies that save power and can communicate over long distances. LoRaWAN is designed for Internet of Things and is suitable for acquiring information from a large number of geographically dispersed sensors [13]. In the agricultural field, it is used to manage sensors that measure cultivation-environment information such as temperature, humidity, and soil moisture and to transmit the sensor information, which contributes to smart agriculture [14]. Smart meters are expected to enhance power resilience and optimization of power transmission and distribution networks, and LoRaWAN is being considered as a network-connection method for these smart meters [15]. An early-detection system for forest fires using LoRaWAN sensor networks has been proposed, which can contribute to the mitigation of forest-fire damages [16]. Thus, LoRaWAN is being implemented or considered for deployment in many fields.

In the field of teleoperation of mobile robots in enclosed spaces, videos are mainly used to investigate the environment. However, LoRaWAN, which has a maximum bit rate of 50 kbps, suffers from the difficulty of transmitting videos. Hence, few discussions are available on the use of LoRaWAN in the field of teleoperation of mobile robots. On the other hand, LoRaWAN is capable of long-distance communication (2–5 km in urban areas and 15 km in suburban areas) [17]. Therefore, LoRaWAN offers the potential for continuous communication even in an environment where wireless LAN (IEEE802.11.g), which has a communication distance of approximately 100 m, is disconnected. Moreover, because LoRaWAN has a frequency bandwidth of 920 MHz, it has high diffusivity and the potential to continue communication even in an enclosed space with many obstacles. Therefore, LoRaWAN is considered to be a suitable wireless-communication standard for subcommunication infrastructures. Thus, in this study, we adopt LoRaWAN as subcommunication infrastructure and explain the teleoperation method that considers the features of LoRaWAN.

## III. TELEOPERATION USING LORAWAN

### A. Existing Teleoperation Protocol using LoRaWAN

In the teleoperation using wireless LAN, the mobile-robot environment is mainly monitored via videos, and the mobile robot is operated using a controller. Because the video is updated at any time, the operator always knows the difference between the current position of the mobile robot and target position and operates the mobile robot to the target position. However, the low bit rate of LoRaWAN makes employing such method of simultaneously learning the environment and moving the mobile robot in teleoperation using LoRaWAN

difficult. Therefore, teleoperation using LoRaWAN requires a separate operation for learning the environment and moving.

In the LoRaWAN communication, a communication path is constructed by designating two communication terminals as a coordinator and an end device. Packets are transmitted from the end device to the coordinator. In the existing teleoperation protocol, the mobile robot transmits information or performs actions in response to commands sent by the operator by considering this feature into account (Fig. 3) [18]. The mobile robot can be operated using such protocol after acquiring the necessary information for teleoperation even in low-bit-rate LoRaWAN. In the next section, we describe the environment learning and operation methods for teleoperation using LoRaWAN.

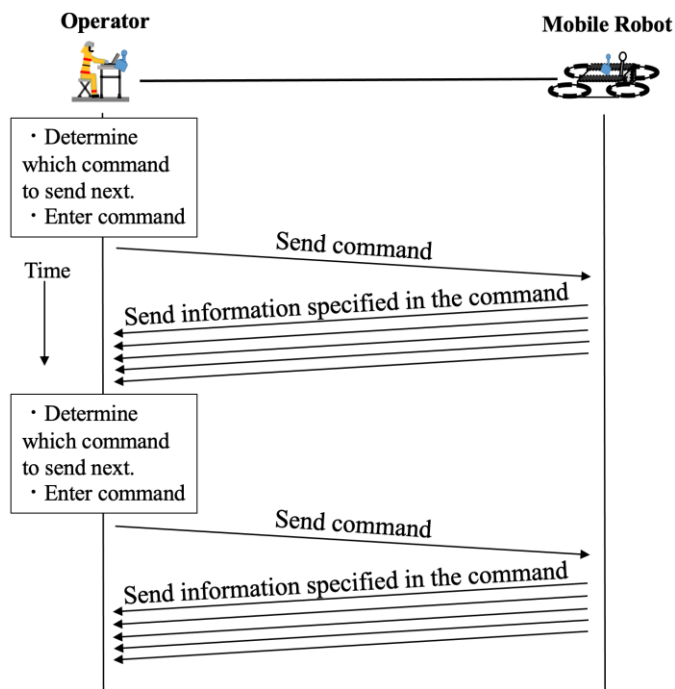


Fig. 3. Teleoperation protocol using LoRaWAN considering a one-direction communication

**B. Related Work**

When teleoperating a mobile robot, learning the surrounding is necessary to confirm the obstacles around the mobile robot and determine its movement path. Therefore, accurately learning the environment in teleoperation using LoRaWAN is necessary. In LoRaWAN, a method for learning the environment using images has been proposed [19]. In this method, the obtained image is transmitted in segments, and the receiver reconstructs the received image elements to obtain the original image.

A teleoperation method using a graphical user interface (GUI) has been proposed to move a mobile robot to a target position (Fig. 4) [20]. In this method, the environment is learned based on the obstacle information obtained by LRF, in addition to the images. The black squares at the interface indicate the presence of obstacles, whereas the ochre squares indicate the absence of obstacles and possibility of movement. Then, when the mobile robot is operated using GUI, its

movement path is determined by sequentially selecting the adjacent ochre-colored cells from the cells where the mobile robot is located. However, this method assumes movement on a level ground with obstacles and cannot handle movement in a 3D space such as a staircase. Therefore, if a distressed mobile robot is required to ascend or descend stairs to evacuate to a place where wireless LAN communication is possible, it cannot achieve its purpose. Hence, we need to consider the process of ascending and descending stairs by the mobile robot using LoRaWAN teleoperation. Section C describes the mobile robot used for stair climbing and descending, and Section D describes the teleoperation method in the considered stair environment.

**C. Mobile Robots Used in Staircase Environment**

Mobile robots with a flipper arm, which is a crawler mechanism (crawler robots), show higher operating performance on uneven terrain such as stairs than wheeled mobile robots. Therefore, in dealing with the Fukushima Daiichi Nuclear Power Plant accident, crawler robots such as the PackBot, Quince, and Survey Runner, which were introduced to collect information in the buildings, were required to climb and descend stairs [21-22]. On the basis of this information, the present study investigates the process of ascending and descending stairs using crawler robots.

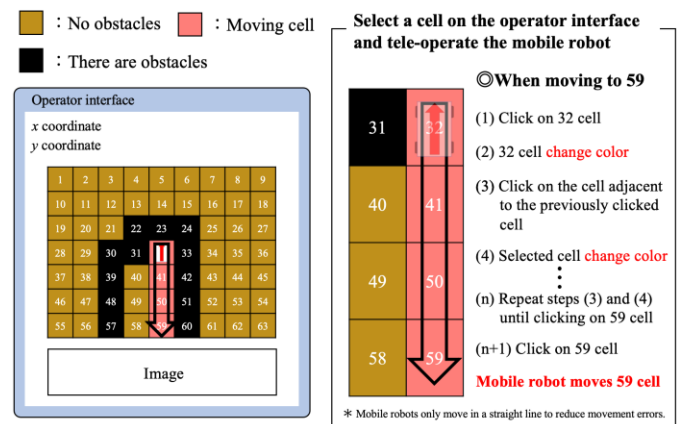


Fig. 4. GUI for remote control using LoRaWAN that considers remote control under a low transmission-capacity communication.

**D. Command-Input Operation Method**

The crawler robot moves and turns by rotating the main crawler and adjusting its posture using its flipper arm. Therefore, in order that the crawler robot can be teleoperated, the operator must be able to send an arbitrary operation command to the main crawler and flipper arm. In addition, for the LoRaWAN teleoperation, moving the crawler robot to the target position based on the acquired information is necessary. The low bit rate of LoRaWAN makes video transmission difficult. Therefore, the current position of the crawler robot as it moves cannot be known in real time. In such an environment, the operator does not know how far the crawler robot has moved if the controller is used in the same manner as that in operation using wireless LAN. Therefore, teleoperating the crawler robot to its target position using a controller is difficult.

Hence, we propose a command-input operation method as a teleoperation method for crawler robots in a staircase



environment. In this method, the operator and crawler robot share a correspondence table of actions, as shown in Fig. 5, and the operator operates the crawler robot by transmitting the numbers in the table. The amount of performed selected operations is determined by specifying the amount of movements such as distance or angle. Thus, the operator can teleoperate the crawler robot to the target position and posture using this method even in an environment where the operator does not learn the change in the current position of the crawler robot in real time. In this study, we select the information required by the operator as follows, i.e., factors (1)–(5), to use the command-input operation method.

- 1) Multiple camera images
- 2) IMU sensor information
- 3) Staircase information (slope, kick-up width, and tread)
- 4) Distance information between the crawler robot and stairs
- 5) Obstacles and staircase-shape information

Item (1) is needed to confirm a safe place where the crawler robot can move forward by learning the stair environment such as the presence or absence of obstacles on the staircase and missing staircase parts. Item (2) is necessary to obtain the posture of the crawler robot in the staircase environment and determine the degree of movement to the target posture. Items (3) and (4) are necessary for determining the amount of movement to move the crawler robot to the target position. Item (5) denotes a factor that determines whether the staircase contains obstacles or part of the staircase is missing. If so, it determines whether these obstacles affect the crawler-robot operation.

No.	Before	After	No.	Before	After
1 Forward			2 Backward		
3 Forward Pivot Left			4 Forward Pivot Right		
5 Spin Left			6 Spin Right		
7 Rear FP up <small>FP : Flipper Arm</small>			8 Rear FP down		
9 Front FP up			10 Front FP down		
11 Rear FP Left up			12 Rear FP Right up		
13 Rear FP Left down			14 Rear FP Right down		
15 Front FP Left up			16 Front FP Right up		
17 Front FP Left down			18 Front FP Right down		

Fig. 5. Action numbers and crawler-robot behavior shared between the operator and crawler robot

#### IV. TELEOPERATION PROTOCOL USING LoRAWAN

##### A. Problem of Teleoperation using LoRaWAN in a Staircase Environment

In a disaster-stricken enclosed space, obstacles exist, which are assumed to block the path. If an obstacle exists in the crawler-robot direction of movement during the evacuation action, it needs to run through the obstacle by climbing over it or by avoiding it using an obstacle-free path.

If obstacles are present in the stair environment, the slope of the stair changes. If the crawler robot tries to avoid these obstacles by climbing over it, it may fall down due to its insufficient climbing ability and very steep gradient when descending (Fig. 6). In addition, if the crawler robot rides on an unstable obstacle, a risk exists in that the crawler robot may fall over because of collapse of the obstacle. Moreover, if the crawler robot tries to avoid the staircase by diverting to an unobstructed area, it may lose its balance and fall down when it turns on the staircase. Thus, a risk exists in which the crawler robot may fail to ascend or descend stairs if obstacles are present in the stair environment.

The disaster-stricken environment assumed in this study is subject to rapid environmental changes and sudden occurrence of obstacles. In addition, the transmission of various sensor information required to send a single operation command takes several minutes. These factors indicate that a risk of failure in the evacuation actions exist because of the occurrence of sudden obstacles in the staircase environment during climbing and descending stairs by teleoperating using LoRaWAN. Therefore, for successful evacuation in staircases with risk of sudden occurrence of obstacles, teleoperation using LoRaWAN is required to shorten the time needed for climbing or descending stairs and completing the stair-climbing process before obstacles appear.

If the amount of transmitted information is limited, e.g., by limiting the number of transmitted images, the teleoperation time can be shortened, but a risk exist in not being able to accurately learn the environment. Therefore, when the teleoperation time is reduced, we must not reduce the information-transmission time during teleoperation but increase the ratio of the time when information is transmitted by reducing the time when no information is transmitted.

As presented in Section III, in the existing teleoperation protocol that uses LoRaWAN, the crawler robot transmits information or performs action in response to a command transmitted by the operator. Therefore, the operator that teleoperates the crawler robot repeatedly decides the command, inputs the command, and receives data. In the existing protocol, no communication occurs when the operator decides which command to send next and when the command is input (bandwidth is not used). Under the same transmission-capacity environment, the amount of data transmission depends on the bandwidth utilization. Therefore, the time required to obtain the information necessary for operation increases with decreasing bandwidth utilization and decreases with increasing bandwidth utilization. On the basis of this discussion, the proposed protocol needs to reduce the teleoperation time by improving the bandwidth utilization.

In the existing protocol, when the operator transmits an incorrect command, the operator cannot interfere and must wait until the crawler robot finishes executing the issued command. Therefore, the teleoperation time is increased by the execution of the wrong command. Hence, the proposed protocol must be able to deal with wrong command execution.

In this paper, we propose a teleoperation protocol that reduces the teleoperation time by taking into account what has been described in this section.

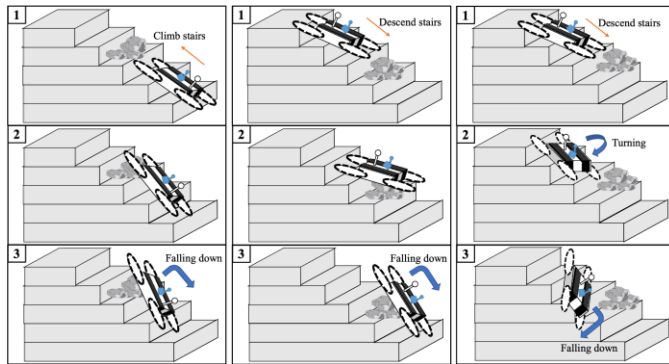


Fig. 6. Crawler robot ascending or descending stairs when obstacles are present on stairs

### B. Interventional Teleoperation Protocol

In this paper, we propose an interventional teleoperation protocol as a teleoperation protocol that satisfies the required specifications. The proposed protocol reduces the command input time and improves bandwidth utilization by sequentially transmitting information necessary for operation according to a predetermined order, instead of transmitting information according to the command received from the operator. However, in this communication flow, even if sufficient information is collected to determine the operation command, transmission is not interrupted, and we need to wait until all the information transmission is completed. Therefore, this communication flow may increase the bandwidth utilization but may not reduce the teleoperation time. Hence, in the proposed protocol, information transmission can be interrupted at arbitrary times using an interventional command, and operation commands can be executed. In addition, the operator can also interrupt the execution of a wrong command using an interventional command.

In the interventional teleoperation protocol, the system must be able to transmit packets from the operator to the crawler robot while it transmits packets to the operator. Therefore, when this method is employed, we assume that two communication paths exist: “operator→crawler robot” and “crawler robot→operator.”

The proposed interventional teleoperation protocol is shown in Fig. 7, and its flow and processing are described hereunder.

1) The operator transmits a synchronized flag (SYN) packet to the crawler robot.

2) The crawler robot sends back an acknowledgement flag (ACK) packet as an acknowledgement of the SYN packet to

the address from which the SYN packet was sent. The crawler robot also transmits SYN packets to verify that packets from the crawler robot reach the operator.

3) The operator confirms the communication connection by receiving an ACK packet and returns the ACK packet in response to a SYN packet sent by the crawler robot.

4) The crawler robot sequentially transmits the sensor information acquired by it after receiving ACK in (3). The order of transmission is concrete numerical information (shape information of obstacles and stairs, IMU sensor information, and distance information between the crawler robot and stairs) and images. In the process of ascending or descending stairs by the crawler robot, we assume that multiple images such as subjective and bird's-eye-view images are used, and the order in which these multiple images are transmitted is set in advance.

5) When a task with a higher priority occurs during the image transmission, an interventional command is sent, the current task is interrupted, and the contents of the command are executed. For example, during image transmission, changing the transmission order of predefined images may be necessary, some images may be lost due to packet loss, or learning of the environment may be completed without waiting for all images to be acquired. Then, when the transmission order of the images is changed, the order is changed by transmitting a new image-transmission order as an interventional command. If the priority of retransmission of a packet-lost image is high, the interventional command to execute the retransmission process is transmitted, and the current image transmission is interrupted to execute the command contents. When information gathering is completed, the operation command is transmitted as an interventional command to operate the crawler robot. When the crawler robot has completed executing the interventional commands other than the operation command, it resumes the task it performed before the intervention command was sent.

When a wrong command is executed, its execution can be interrupted by transmitting the interventional command. Thus, the interventional command can be used to reduce the time used in the bandwidth and to deal with the execution of wrong command contents.

6) When the crawler robot receives an operation command as an interventional command in (5), it interrupts the image transmission and executes the command. When the crawler robot moves, highly real-time information such as IMU sensor or encoder information is always transmitted to the operator.

7) The operator checks whether the crawler robot has moved more than the specified distance based on the information obtained in (6) and transmits an interventional command to stop the crawler robot if necessary.

V. EVALUATION EXPERIMENT

A. Purpose and Content of the Experiment

The purpose of this experiment is to confirm whether the proposed method can reduce the teleoperation time of the crawler robot. In our experiments, we measure the bandwidth utilization and teleoperation time of the crawler robot by conducting stair climbing and descending using the proposed and existing teleoperation protocols. Bandwidth utilization is defined as the ratio of the time the crawler robot transmits packets to the operator during the teleoperation time. Teleoperation time is defined as the time from that when the crawler robot transmits the first packet to the operator to the time when the crawler robot completes ascending or descending the stairs and confirms that it is at the floor or landing level.

B. Experimental Environment and Equipment

The experimental environment is shown in Fig. 9. In the experiment, LRF, which was mounted in front of the crawler robot, was used to acquire information on the shape of the obstacles and stairs as well as the distance between the crawler robot and stairs. Three cameras were used, which were positioned to acquire subjective, overhead, and rearward images. The control system used in this experiment was Raspberry Pi 4 Model B, and the LoRaWAN module was ES920LR manufactured by EASEL Corporation.

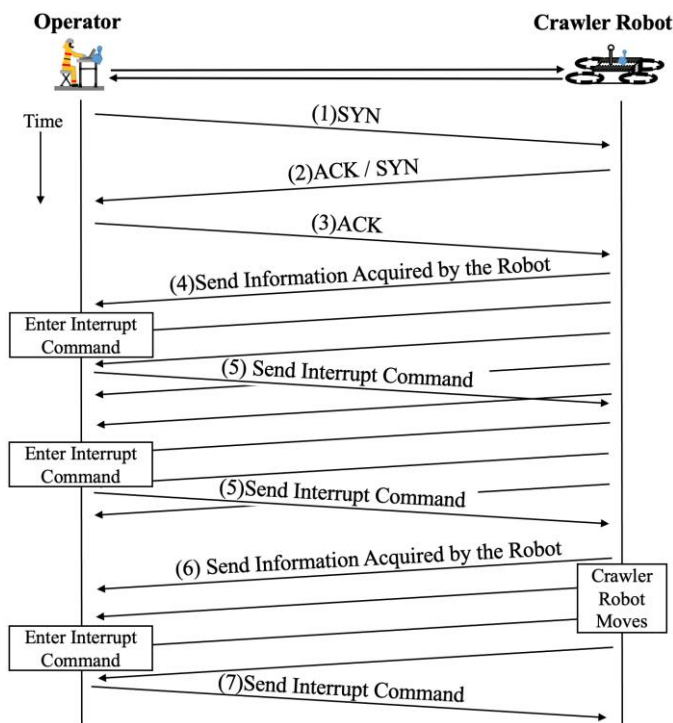


Fig. 7. Interventional teleoperation protocol

The packet format used in the proposed protocol is shown in Fig. 8. The interventional command that is part of the packet used in (5) contains commands to “change the order of image transmission,” “retransmission processing,” “image-quality conversion,” and “cancellation of the last sent command”.

The packet format used in (4)–(6) contains a packet number at the end. This packet number is a number expressed in ascending order starting from zero, and the operator can understand the order of packets that the crawler robot have sent. When the order of packets transmitted by the crawler robot and that received by the operator are different, the operator can learn the difference using these packet numbers. In addition, if the same command is delivered to the crawler robot more than once, the packet number can be used to cancel multiple executions of the same command.

Header Fields					1 Byte
Frame Control	PAN ID	Destination Address	Sender Address	Communication Route	SYN/ACK

Packet used to send SYN/ACK packet in (1), (2), and (3)

Header Fields					Variable	1 Byte
Frame Control	PAN ID	Destination Address	Sender Address	Communication Route	Sensor Information	Packet Number

Packet used to send sensor information in (4), (6)

Header Fields					1 Byte	Variable	1 Byte
Frame Control	PAN ID	Destination Address	Sender Address	Communication Route	Interruption Command	Detail of Command	Packet Number

Packet used to send interruption command in (5)

Header Fields					1 Byte	1 Byte	1 Byte
Frame Control	PAN ID	Destination Address	Sender Address	Communication Route	Operation Command	Amount of Movement	Packet Number

Packet used to send operation command in (5), (7)

Fig. 8. Packet format used in the proposed protocol

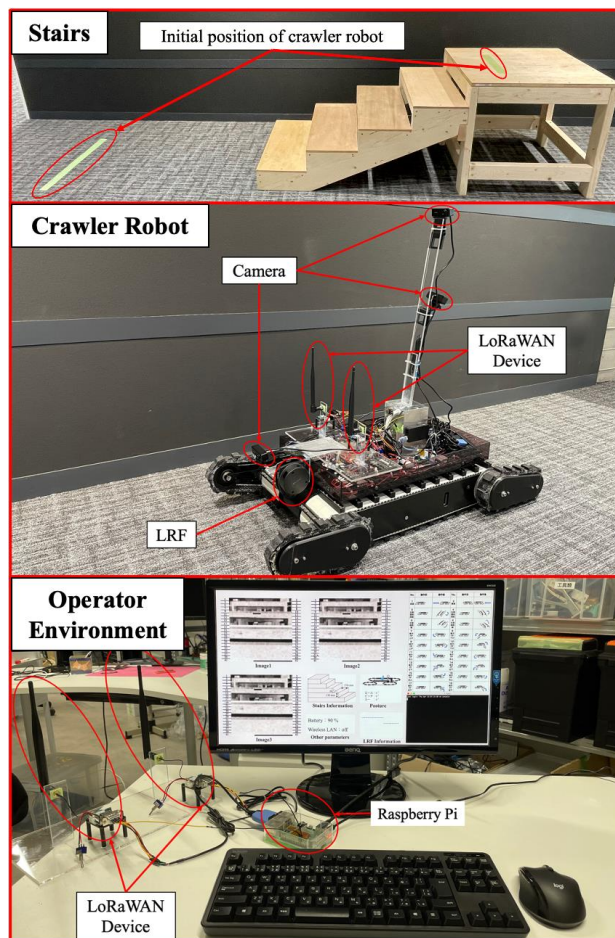


Fig. 9. Experimental environment

### C. Experimental Result

In the experiment, the crawler robot was able to ascend and descend stairs without tipping over once in both cases. Fig. 10 shows the experimental results that used the existing method, and Fig. 11 shows those that used the proposed method. The vertical axis indicates whether the bandwidth was used or not, where one denotes that the bandwidth was used and zero indicates that the bandwidth was not used.

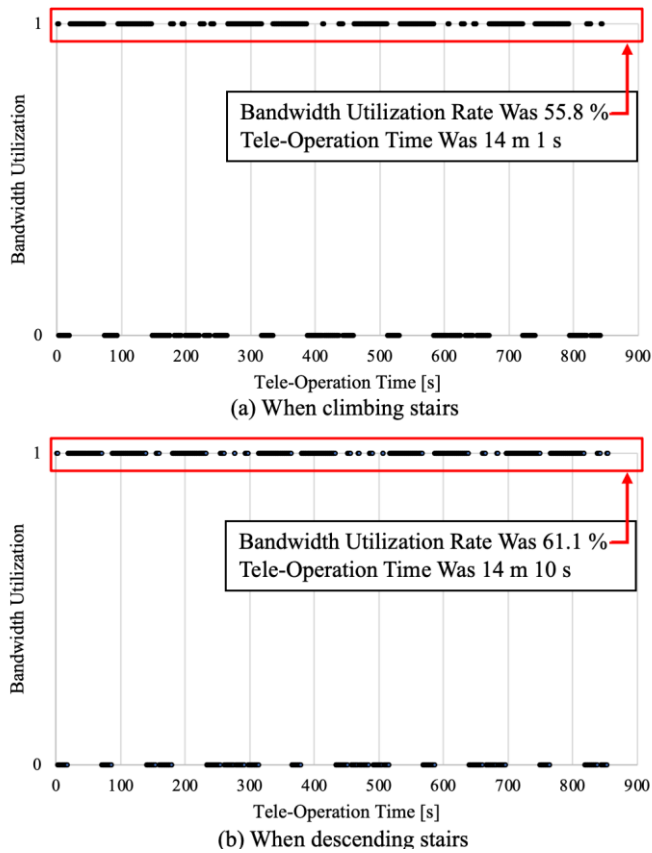


Fig. 10. Results of the experiment (existing protocol)

In the ascent process using the existing method, concrete numerical information (IMU sensor information, distance between the crawler robot and stairs, and shape information of the obstacles and stairs) was transmitted seven times, images were transmitted eight times, and the crawler robot moved six times. In the case where the existing method was used, concrete numerical information was transmitted six times, images were transmitted nine times, and the crawler robot was moved seven times. In the case of the ascent process using the proposed method, concrete numerical information was transmitted seven times, images were transmitted ten times, and the crawler robot was moved six times. In the case of the descent process using the proposed method, concrete numerical information was transmitted nine times, images were transmitted ten times, and the crawler robot moved eight times.

In the ascent process, the concrete numerical information and number of times the crawler robot moved were the same in both methods, but the number of images acquired by the proposed method increased by two. The bandwidth utilization improved by 38.3% from 55.8% to 94.1%, and the

teleoperation time decreased by 2 min and 9 s from 14 min and 1 s to 11 min and 52 s. In the descent process, the amount of concrete numerical information acquired using the proposed method was increased by three, the number of images was increased by one, and the number of crawler-robot movements was increased by one. The bandwidth utilization improved by 32.5% from 61.1% to 93.6%, and the teleoperation time decreased by 1 min 58 s from 14 min 10 s to 12 min 12 s.

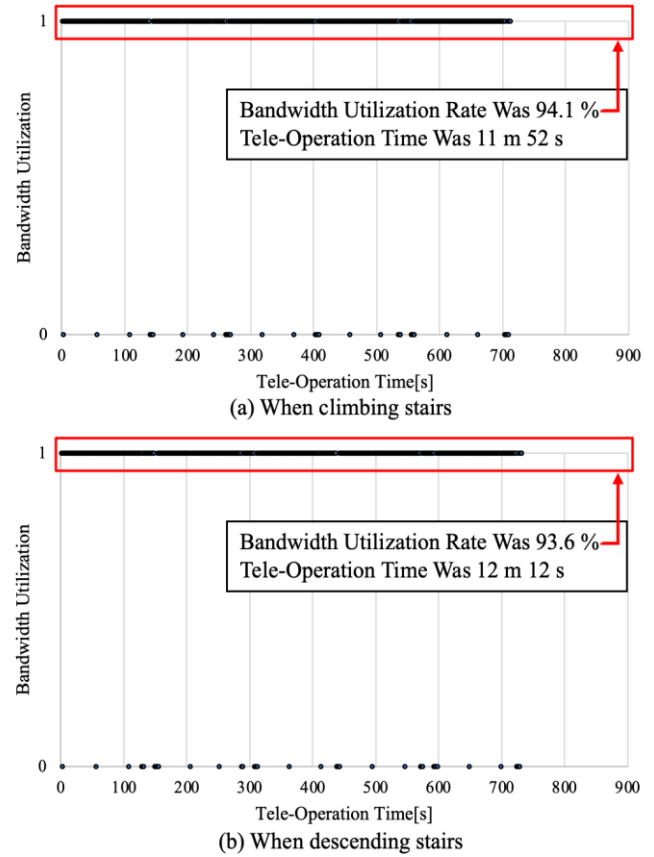


Fig. 11. Results of the experiment (proposed protocol)

The average time required to determine and input the next command using the existing method was 19.6 s for ascending and 15.7 s for descending. In the proposed method, the average time from the completion of obtaining the necessary information to inputting of the operation command was 21.8 s for ascending and 17.0 s for descending.

### VI. DISCUSSION

We confirmed that the proposed method could reduce the teleoperation time, although two more images were obtained in addition to the information acquired by the existing method when the crawler robot was ascending. We confirmed that the proposed method could shorten the teleoperation time, although three times of concrete numerical information and one more image were obtained in addition to the information acquired by the existing method when the crawler robot was descending. Furthermore, the average time of the proposed method from acquisition of the necessary information to inputting of the operation command was longer than the average time required for the decision and inputting of the next

command in the existing method, both in the stair climbing and descending processes. Therefore, the reduction in the teleoperation time in the experiment did not depend on the time to decide and input the commands. Hence, the teleoperation time could be reduced by reducing the number of command inputs and improving the bandwidth utilization using interventional commands. Therefore, this method is considered to be effective for teleoperation in staircases with a large number of command inputs because of the need for concrete numerical information and multiple images.

The teleoperation that uses the proposed method could interrupt the execution of a command sent by mistake because we confirmed that the interventional process could be executed. Therefore, this method is considered to be effective because it could deal with increased teleoperation time when wrong command contents were executed.

## VII. CONCLUSION

In the teleoperation of crawler robots in a disaster-stricken enclosed space, the distress of the crawler robot due to communication breakdown is a problem. In this study, we have presented a teleoperation method using LoRaWAN as a subcommunication infrastructure to move a distressed crawler robot to the place where wireless LAN communication is possible. In this study, we have also assumed an environment in which the crawler robot was required to ascend and descend stairs when evacuating to the place where wireless LAN communication was possible. We have presented the operation method of the crawler robot on stairs.

In the disaster-stricken environment assumed in this study, sudden appearance of obstacles is expected. When obstacles appear on the staircase, the slope of the staircase changes and the crawler robot may fail to ascend or descend the stairs because of its insufficient climbing ability. In addition, because of the low bit rate of LoRaWAN, several minutes is required to obtain the necessary information to transmit a single operation command. Hence, a risk exists in terms of failure in evacuation actions due to the occurrence of obstacles during stair climbing or descending teleoperation using LoRaWAN. Therefore, shortening the time required for stair climbing or descending and completing the stair climbing or descending process are important before obstacles appear in the teleoperation using LoRaWAN. Hence, we propose an interventional teleoperation protocol to reduce the teleoperation time.

The proposed method reduces the teleoperation time by reducing the number of command decisions and inputs and employing an interventional command. In addition, the proposed method can interrupt the execution of erroneous commands using interventional commands. Through experiments in real environments, we confirmed that the proposed method could improve the bandwidth utilization and shorten the teleoperation time.

In this study, we developed a multipath communication method that increased the bandwidth by adding more communication paths. Therefore, we plan to increase the transmission capacity of the bandwidth by adding a communication path from the crawler robot to the operator to further shorten the teleoperation time.

## REFERENCES

- [1] Yoshiaki Kawata, "The great Hanshin-Awaji earthquake disaster, damage, social response, and recovery," *Journal of Natural Disaster Science*, Vol. 17, No. 2, pp. 1-12, 1995.
- [2] Jingxuan Sun, Boyang Li, Yifan Jiang, Chih-yung Wen, "A camera-based target detection and positioning UAV system for search and rescue (SAR) Purposes," *Sensors*, Vol. 16, No. 11, pp. 1778, 2016.
- [3] F. Kurz, D. Rosenbaum, J. Leitloff, O. Meynberg, P. Reinartz, "A real time camera system for disaster and traffic monitoring," *Proceedings of International Conference on SMPR*, pp. 1-6, 2011.
- [4] Sabarish Chakkath, "Mobile robot in coal mine disaster surveillance," *IOSR Journal of Engineering*, Vol. 2, No. 10, pp. 77-82, 2012.
- [5] Masataka Fuchida, Shota Chikushi, Alessandro Moro, Atsushi Yamashita, Hajime Asama, "Arbitrary viewpoint visualization for teleoperation of disaster response robots," *Journal of Advanced Simulation in Science and Engineering*, Vol. 6, No. 1, pp. 249-259, 2019.
- [6] T. B. Bhondve, R. Satyanarayan, M. Mukhedkar, "Mobile rescue robot for human body detection in rescue operation of disaster," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, Vol. 3, No. 6, pp. 9876-9882, 2014.
- [7] H. A. Reddy, B. Kalyan, Ch. S. N. Murthy, "Mine rescue robot system—a review," *Procedia Earth and Planetary Science*, Vol. 11, pp. 457-462, 2015.
- [8] Tomoaki Yoshida, Keiji Nagatani, Satoshi Tadokoro, Takeshi Nishimura, Eiji Koyanagi, "Improvements to the rescue robot Quince toward future indoor surveillance missions in the Fukushima Daiichi nuclear power plant," *Field and Service Robotics*, Vol. 92, pp. 19-32, 2013.
- [9] K. Nagatani, S. Kiribayashi, Y. Okada, S. Tadokoro, T. Nishimura, T. Yoshida, E. Koyanagi, Y. Hada, "Redesign of rescue mobile robot Quince," *9th IEEE International Symposium of Safety, Security, and Rescue Robotics*, pp. 13-18, 2011.
- [10] Joshua Reich, Elizabeth Sklar, "Robot-sensor networks for search and rescue," *Proceedings of the IEEE*, 2007.
- [11] Yuta Koike, Kei Sawai, Tsuyoshi Suzuki, "A study of routing path decision method using mobile robot based on distance between sensor nodes," *International Journal of Advanced Research in Artificial Intelligence*, Vol. 3, No. 3, pp. 25-31, 2014.
- [12] Tsuyoshi Suzuki, Ryuji Sugizaki, Kuniaki Kawabata, Yasushi Hada, Yoshito Tobe, "Autonomous deployment and restoration of sensor network using mobile robots," *International Journal of Advanced Robotic Systems*, Vol. 7, No. 2, pp. 105-114, 2010.
- [13] Alexandru-Ioan Pop, Usman Raza, Parag Kulkarni, Mahesh Sooriyabandara, "Does bidirectional traffic do more harm than good in LoRaWAN based LPWA networks?," *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1-6, 2017.
- [14] Danco Davcev, Kosta Mitreski, Stefan Trajkovic, Viktor Nikolovski, Nikola Koteli, "IoT agriculture system based on LoRaWAN," *2018 14th IEEE International Workshop on Factory Communication Systems*, pp. 1-4, 2018.
- [15] Alvin Yusri, Muhammad Iman Nashiruddin, "LORAWAN internet of things network planning for smart metering services," *2020 8th International Conference on Information and Communication technology*, pp. 1-6, 2020.
- [16] Georgi Hristov, Jordan Raychev, Diyana Kinaneva, Plamen Zahariev, "Emerging methods for early detection of forest fires using unmanned aerial vehicles and LoRaWAN sensor networks," *2018 28th EAAEIE Annual Conference*, pp. 1-9, 2018.
- [17] Ferran Adelantado, Xavier Vilajosana, Pere Tuset-Peiro, Borja Martinez, Joan Melia-Segui, Thomas Watteyne, "Understanding the limits of LoRaWAN," *IEEE Communication Magazine*, Vol. 55, No. 9, pp. 34-40, 2017.
- [18] Toshihiro Yamasaki, Kei Sawai, Noboru Takagi, Tatsuo Motoyoshi, Hiroyuki Masuta, Kenichi Koyanagi, "Development of routing method considering multi-hop using LoRaWAN as sub-communicator for mobile robot," *SICE SI2019*, 3E1-09, 2019. (In Japanese)

- [19] Akram Jebril, Aduwati Sali, Alyani Ismail, Mohd Fadlee Rasid, "Overcoming limitations of LoRa physical layer in image transmission," *Sensors*, Vol. 18, No. 10, pp. 3257, 2018.
- [20] Toshihiro Yamasaki, Tatsuo Motoyoshi, Hiroyuki Masuta, Noboru Takagi, Kei Sawai, "Development of evacuation operation method for mobile robot in low transmission rate communication using LoRaWAN," *SICE SI2020*, 1E2-10, 2020. (In Japanese)
- [21] Keiji Nagatani, Takeshi Nishimura, Yasushi Hada, Yoshito Okada, Satoshi Tadokoro, Tomoaki Yoshida, Eiji Koyanagi, "Redesign of rescue mobile robot Quince," *IEEE International Symposium on Safety, Security, and Rescue Robotics*, pp. 13-18, 2011.
- [22] Qihao Zhang, Wei Zhao, Shengnan Chu, Lei Wang, Jun Fu, Jiangrong Yang, Bo Gao, "Research progress of nuclear emergency response robot," *IOP Conference Series: Materials Science and Engineering*, Vol. 452, No. 4, pp. 042102, 2018.

# Business Intelligence Data Visualization for Diabetes Health Prediction

## Data Analytics and Insights for Diabetes Prediction

Samantha Siow Jia Qi, Sarasvathi Nagalingham

Faculty of Data Science and Information Technology  
INTI International University, Nilai, Negeri Sembilan, Malaysia

**Abstract**—In today's environment, Business Intelligence (BI) is transforming the world at a rapid pace across domains. Business intelligence has been around for a long time, but when combined with technology, the results are astounding. BI is also playing an important role in the healthcare domain. Centers for Disease Control and Prevention (CDC) is the largest science-based, data-driven service provider in the country for public health protection. For over 70 years, has been using science to fight disease and keep families, businesses, and communities healthy. However, research indicates that the prevalence of diabetes in the US is rising alarmingly. As a result, if diabetes is not treated, it can lead to life-threatening complications such as heart disease, loss of feeling, blindness, kidney failure, and amputations. As a result, this study was conducted to analyze people's health conditions and daily lifestyles in order to predict which type of diabetes they would most likely diagnose with the implementation of business intelligence using Tableau dashboard. Furthermore, background research is conducted on CDC to understand their work, challenges, and opportunities. By the end of the project, the information obtained and visualized should be able to enhance business choices and make better decisions on controlling diabetes in the future.

**Keywords**—Diabetes; business intelligence; prediction; dashboard visualization; data analysis; centers for disease control and prevention

### I. INTRODUCTION

#### A. Business Intelligence Methodology

Healthcare analytics and business intelligence (BI) are two emerging technologies that offer analytical capability to aid the healthcare sector in enhancing service quality, lowering expenses, and managing risks. With the increasing volume of data and the desire to learn from it, demand for BI applications for healthcare continues to rise. The need for data management and analysis expertise in healthcare is rapidly increasing. With the help of Business Intelligence, companies can make better decisions by displaying current and historical data within the context of their business. Analysts may use BI to generate the company's performance and competitive standards, which will help the companies operate more smoothly and effectively as well as identify the market trends for enhancement. In any BI project, a consistent methodology and approach must be defined as they can let decision-makers develop, support, and integrate best management practices within the organization. Moreover, it can achieve a higher chance of success, save time and effort, eliminate unnecessary

operations, and ensure accurate reporting and analysis. Therefore, the Agile BI methodology will be adopted in the context of our domain, Centers for Disease Control and Prevention (CDC).

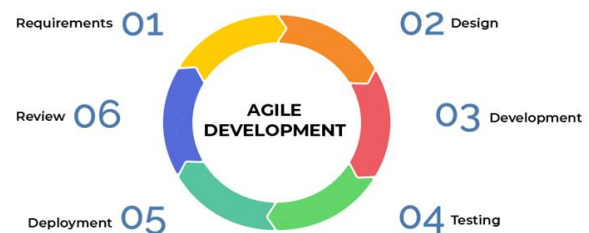


Fig. 1. Phases of agile methodology.

The Agile BI methodology is a project management strategy that focuses on continuous improvement and is suitable for those that demand speed and flexibility to satisfy customer requirements. It is considerably more accessible than it looks to apply Agile BI methodology to an organization's internal business processes [5]-[6]. The following are the phases of the developing an Agile BI methodology (Fig. 1):

- **Requirement:** In this phase, the project team specifies the requirements. They should outline the business opportunities of the project that can result in profit and business growth, along with estimating the time and effort required to complete it. Technical and economic feasibility can be determined as well as whether the project is worth pursuing based on this information.
- **Design:** This phase required working with the company's stakeholders to determine the requirements once the project had been identified. Diagrams such as flow diagrams and high-level UML diagrams are used to demonstrate how the new features perform and how they integrate into the existing system.
- **Development:** The works begin once the project team has identified the requirements based on stakeholder feedback. UX designers and developers start working on the project's initial iteration to deliver a viable product. The product will go through several phases of improvement to meet the stakeholder's requirements.

- **Testing:** The Quality Assurance team will check the product's performance and address any defects and bugs that produce unexpected results. The software engineer team will finalize the system and user documentation once the product is satisfied.
- **Deployment:** This phase includes ongoing software release support, especially when the project team must keep the system working well and demonstrate users on using the system. The production phase comes to an end when the support has ended.
- **Review:** The final phase is to get feedback from the stakeholder once the product has been released. The project team will receive feedback on the development and work through it to improve their work for the next project.

### B. Literature Review

The Centers for Disease Control and Prevention (CDC) is the United States' main public health agency. It is a government agency of the United States that belongs to the Department of Health and Human Services and is headquartered in Atlanta, Georgia. The agency's primary aim is to protect public health and safety by controlling and preventing disease, injury, and disability in the United States and around the world. The CDC concentrates national attention on developing and implementing disease control and prevention strategies [7].

The Behavioral Risk Factor Surveillance System (BRFSS) is a health-related telephone survey conducted annually by the Centers for Disease Control and Prevention (CDC). The survey collects data from over 400,000 Americans each year on health-related risk behaviors, chronic health issues, and use of preventative treatments. It has been conducted every year since 1984. According to the CDC, 34.2 million Americans have diabetes and 88 million have prediabetes as of 2018. Furthermore, the CDC indicates that one in every five diabetics and about eight out of ten pre diabetics are unaware of their risk. To improve preventive care, EHR is necessary, which can monitor the patient's state of health and track treatment progress to make preventative care easier.

The main purpose of this business is to implement electronic health records (EHR) to give better care to patients. EHR and the ability to exchange health information electronically can help patients receive better quality and safer care while also providing substantial profit to the business. It automates a multitude of operations for the practice and allows for instant access to patient records for better coordinated and efficient care when compared to manually inserting, updating, or deleting data using a csv file. To improve productivity and work-life balance, physicians are allowed to share electronic information with one another remotely and in real time [1].

Furthermore, early diagnosis is essential since it can lead to lifestyle changes and more effective treatment, making diabetes risk prediction models important tools for the general population and public health officials. Health analytics are given by applying EHR, which can help to recognize patterns, predict diagnoses, and recommend potential treatment

alternatives. Rather of relying on trial-and-error methods, these analytics lead to more successful overall patient results the first time.

### C. Challenges faced by CDC

Various changes are occurring in the Centers for Disease Control and Prevention (CDC), creating new challenges to both large and small medical businesses. To name a few, the integration of services, technological developments, and patient preferences have created a new environment in which performing a medical treatment is no longer solely about treating patients.

First, big data development is well received in the medical world, but implementation may not be as easy. Non-relational databases combine patient data from several sources to provide meaningful metrics. The technology appears at the ideal time to meet the demand for newly available patient information warehouses. Relational databases have traditionally been used by healthcare providers to manage and store patient records. Nevertheless, relational databases are incapable of managing unstructured data, such as medical documentation and transcripts. Only a small fraction of healthcare providers has successfully transitioned from relational to non-relational databases using standard electronic health records (EHRs). Most firms that effectively employ non-relational information systems are big and financially secure.

Besides that protecting the devices that protect public health is important as many developments in healthcare technology use Internet connectivity. However, this convenience extends to cyber attackers as well. Malicious cyber-attacks will become more common as the Internet-connected medical device industry evolves. The US authorities warned people in 2015 that hackers may command infusion pumps to deliver harmful drug doses. The leak raised the prospect of malevolent programmers infiltrating medical devices and causing harm to people. Moreover, hackers utilize medical equipment to invade care provider networks, stealing research and clinical trial data as well [13]. Presently, no medical equipment has been infiltrated by hackers, resulting in a fatal event. Yet, cyber security experts warn that an assault on an untrained care provider can harm an organization in a variety of ways.

Furthermore, as patients are becoming more liable for a growing amount of their medical expenses, health providers are listing patient collections as their top revenue cycle management challenge. Providers must comply with patient payment preferences to urge patients to make payments on time. Invoice statements should be patient-friendly to meet the expectations of patients and enhance their user experience. For example, e-Statements and a range of payment alternatives, such as credit cards, etc. through an online patient portal can be used. However, setting up such billing and payment processing systems in-house can be difficult and expensive for medical offices. They must not only discuss agreements with each payment processor and develop the infrastructure, but they must also bear the continuous administrative costs of such technology. Besides, healthcare professionals must adhere to stringent criteria to preserve patient information.



You must ensure that your payment interface and processing system are entirely compliant, or you may face a severe penalty.

Lastly, in recent years, there have been substantial changes in the health insurance sector of the CDC. As more patients bear a greater share of their medical costs, they naturally expect greater services from their providers. The CDC will confront increased competition in gaining and maintaining patients who seek a level of customer service comparable to that of other retail chains. For example, they expect a streamlined patient experience in which they can "self-service" to settle most doubts, problems, or concerns whenever, wherever, and however they see fit [2], [14]-[15].

#### D. Opportunities of CDC

After looking into the difficulties or problems that arise in the disease control and prevention centers, we have identified some alternative solutions as well as opportunities to ensure the organization remains sustainable. The opportunities of disease control and prevention centers are listed as follows.

Due to the Centers for Disease Control and Prevention (CDC) currently using the traditional spreadsheet approach to store and manage patient records, it slows down the workflow of the organization. Therefore, CDC needs to transition from using a spreadsheet like Microsoft Excel to an electronic health records (EHR) system. The EHR system helps their CDC staff to capture and access information gathered during patient appointments with just a click. It is more user-friendly than a spreadsheet since the process of adding, updating, viewing, and deleting the patient's data can be done systematically and more simply. Moreover, the EHR system allows multiple staff to have simultaneous access to health records anytime and anywhere. And most significantly, it also enables the staff to share information with other practices suchlike emergency facilities, specialists, and laboratories, to contain all the information clinicians provide for patient care [9]-[10].

Besides that CDC will be able to expand the view of patient care and ensure every patient has gotten the medical services they need. With the use of the EHR system, it has provided a variety of useful functions like the capability to generate the analytical report based on clinical data collected from ongoing patient care. In healthcare, the analytical report is used to detect or predict early signs of patient deterioration and aids in the more definitive patient diagnosis, followed by the appropriate treatment of the identified indication. Accurate preventive measures have been shown to reduce mortality and morbidity rates in diagnosed patients.

Lastly, a high level of security management retains patients as EHR systems provide better security of confidential records. Certain users might be granted varying levels of access to patients' data to ensure that the sensitive files are protected and kept safe. The unauthorized individuals who aim to access the system will be restricted. The probability of valuable records being stolen by theft or hacker is lower. As a result, CDC has gained trust and built loyalty among patients.

## II. PROPOSING BI SOLUTIONS

Predictive analytics is one of the top business intelligence trends for the past two years in a row, but the potential applications extend well beyond business and far into the future. CDC can build a database for predictive analytics tools that would enhance care delivery. This is especially important in the case of individuals who have a complicated medical history and are suffering from various illnesses. The purpose of healthcare internet business intelligence is to assist doctors in making data-driven choices in seconds and improving patient care. The purpose of healthcare internet business intelligence is to assist health care agencies in making data-driven choices in seconds and improving the health of the United States citizens. CDC can also use BI tools to make a predictive diabetes analysis in detail. With these, we can clearly know the health of the United States citizens and identify the key factors about who are getting diseases. Then, CDC can find the best solution to reduce the Americans' diseases. However, new BI solutions and tools would also be able to anticipate who is at risk of diabetes, so it is advised on extra testing or weight control [8].

In addition to predictive analysis, BI tools may assist firms in analysing clinical data such as assessing lab test results, the incidence of unfilled prescriptions, and so on. This allows CDC to track disease entails determining what causes individuals to get ill as well as the most efficient measures to avoid illness. It can also assist local committees in determining which regions require further funding.

Moreover, business intelligence is a relatively recent concept that refers to the gathering and analysis of data to better business operations and strategic planning. This similar paradigm underpins healthcare business intelligence, but the data in issue is patient data obtained through several sources. Using the BI tools can improve the decision making and ensure the data quality [12]. Using BI tools can also quickly generate accurate reports. BI tools make it simpler and easier for businesses to develop and share dashboards and gather information [11]. For example, Tableau includes several ready-made templates for users in the healthcare agency, which aids in installation by allowing firms to easily drill down into their data. Healthcare BI software is a subset of BI software that is aimed specifically at the healthcare sector. These technologies increase the ability of medical experts to examine data obtained from various sources. These sources might include patient files and medical data, but they can also include extra information.

## III. DATASET ANALYSIS

The dataset used in this project is retrieved from National Health and Nutrition Examine Survey (NHANES) by CDC [3]. The author did some analysis on the dataset and found out that the dataset is a health record of diabetes chronic disease representing the United States population of all ages. The dataset consists of attributes describing various blood testing, body mass index (BMI) and other symptoms that could cause diabetes. Moreover, it is also found that the dataset does not contain any individual's personal information as health records are confidential information. Due to the NHANES Data Release and Access Policy, these records are designated

as classified official documents because they include private and personal information. Therefore, it must be given a high level of data protection.

### A. Description of Dataset

This dataset contains around 200,000 records with 22 attributes. The attributes and its data type as well as description are shown below (Table I):

TABLE I. DESCRIPTION OF DATASET

No	Attribute Name	Data Type	Description
1	Diabetes_012	Numerical (Interval Scale)	Disease status 0 = no diabetes 1 = prediabetes 2 = diabetes
2	HighBP	Boolean	Blood pressure 0 = no high BP 1 = high BP
3	HighChol	Boolean	Cholesterol 0 = no high cholesterol 1 = high cholesterol
4	CholCheck	Boolean	Cholesterol check 0 = no cholesterol check in 5 years 1 = yes, having cholesterol check in 5 years
5	BMI	Numeric (Continuous)	Value of body Mass Index based on patient weight and height
6	Smoker	Boolean	Have ever smoked more than 100 cigarettes (5 packs) in a lifetime 0 = no 1 = yes
7	Stroke	Boolean	Have ever experienced a stroke 0 = no 1 = yes
8	HeartDiseaseorAttack	Boolean	Have ever experiencing coronary heart disease (CHD) or myocardial infarction (MI) 0 = no 1 = yes
9	PhysActivity	Boolean	Physical activity within the past 30 days 0 = no 1 = yes
10	Fruits	Boolean	Consume fruit at least one or more times per day 0 = no 1 = yes
11	Veggies	Boolean	Consume vegetables at least one or more times per day 0 = no 1 = yes
12	HvyAlcoholConsump	Boolean	Heavy drinkers • Adult men – having more than 14 drinks per week • Adult women – having more than 7 drinks per week 0 = no 1 = yes
13	AnyHealthcare	Boolean	Have any type of health care coverage (health insurance, prepaid plans like HMO)

			0 = no 1 = yes
14	NoDocbcCost	Boolean	In the past 12 months, have you had a need to see a doctor but couldn't because of the cost? 0 = no 1 = yes
15	GenHlth	Numerical (Interval Scale)	Would you say that your overall health is: Scale 1-5 1 = excellent 2 = very good 3 = good 4 = fair 5 = poor
16	MentHlth	Numerical (Interval Scale)	In the last 30 days, how many days did you have poor mental health, including stress, depression, emotional issues? Scale 1-30 days
17	PhysHlth	Numerical (Interval Scale)	In the last 30 days, how many days did you have poor physical health, including physical illness and injury? Scale 1-30 days
18	DiffWalk	Boolean	Have trouble walking or climbing stairwells? 0 = no 1 = yes
19	Sex	Nominal Scale	Gender 0 = female 1 = male
20	Age	Numerical (Ratio Scale)	Level age category 1 = Age 18 to 24 2 = Age 25 to 29 3 = Age 30 to 34 4 = Age 35 to 39 5 = Age 40 to 44 6 = Age 45 to 49 7 = Age 50 to 54 8 = Age 55 to 59 9 = Age 60 to 64 10 = Age 65 to 69 11 = Age 70 to 74 12 = Age 75 to 79 13 = Age 80 or older
21	Education	Numerical (Ratio Scale)	Education level Scale 1-6 1 = Never attended school or only kindergarten 2 = Grades 1 through 8 (Elementary) 3 = Grades 9 through 11 (Some high school) 4 = Grade 12 or GED (High school graduate) 5 = College 1 year to 3 years (Some college or technical school) 6 = College 4 years or more (College graduate)
22	Income	Numerical (Ratio Scale)	Income Scale 1-8 1 = less than \$10,000 5 = less than \$35,000 8 = \$75,000 or more

### B. Data Cleaning

The practice of correcting or deleting inaccurate, damaged, improperly formatted, duplicate, or incomplete data from a dataset is known as data cleaning. There are numerous ways for data to be duplicated or incorrectly categorized when merging multiple data sources. Even if results and algorithms appear to be correct, they are unreliable if the data is inaccurate. Because the procedures will differ from dataset to dataset, there is no one definitive way to specify the precise phases in the data cleaning process. But it is essential to create a template for your data cleaning procedure so you can be sure you are carrying it out correctly each time. The following techniques demonstrate how the dataset is cleaned by using Microsoft Excel and RStudio.

Microsoft Excel provides a lot of functions to the user and allows the user to show the data analysis results in a variety of ways. The data analysis results can be displayed as charts that emphasize the significant points in the data, and the audience will immediately comprehend what you want to project in the data. Excel allows us to visualize the data to reveal hidden data and identify data patterns. Several features have been picked from the vast array to be used in this project.

- Checking Missing Values

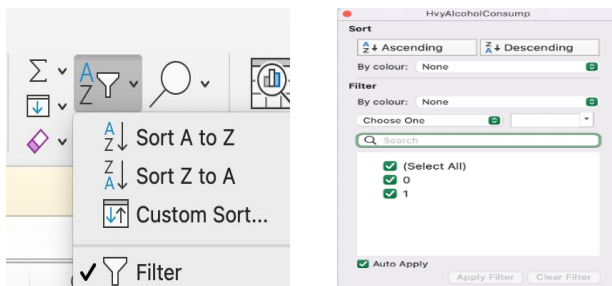


Fig. 2. Filter function in excel.

Filter function is used to filter a range of data depending on the criteria you define (Fig. 2). The filter function allows users to filter out all the data in a column, allowing them to see an overview of the data values. Furthermore, use this method to determine whether there is a null value in the column.

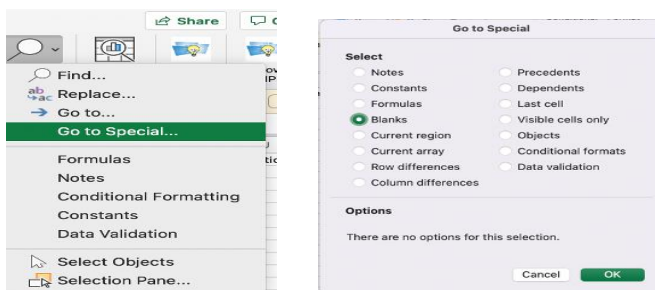


Fig. 3. Find function in excel.

The find function allows the user to check the missing value (Fig. 3). After that put all the missing values as blank. It means the dataset occurs missing value if it manages to be replaced as a blank field. So, it is required to handle it when

there is a null value to ensure data accuracy and improve data quality.

- Remove Duplicate Data

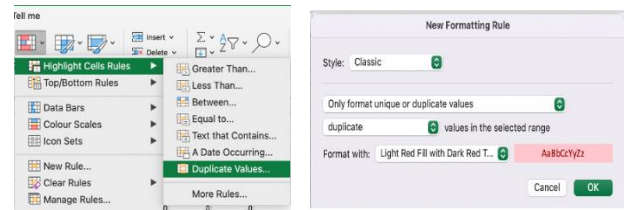


Fig. 4. Remove duplicate function.

Duplicate data might be valuable in some cases, but it can also make it difficult to interpret your data (Fig. 4). To discover and highlight duplicate data, conditional formatting can be used. That way, you may go through the duplicates and determine whether to eliminate them. It allows you to select the formatting you wish to apply to the duplicate values in the box next to values with.

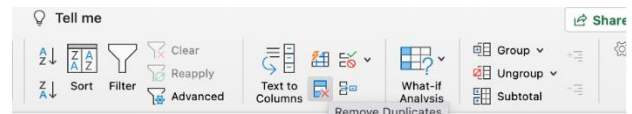


Fig. 5. Remove duplicates feature.

The duplicate data will be permanently removed if you utilize the Remove Duplicates feature (Fig. 5). Before deleting the duplicates, it's a good practice to transfer the original data to another worksheet so you don't lose any information inadvertently. It is also more convenient for users to remove all the duplicate data. It will remove all the duplicate values and remain unique.

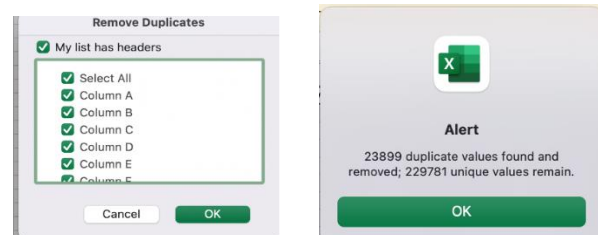


Fig. 6. Remove all duplicates.

As a result, used this method to retrieve the unique value efficiently and quickly from the specified column (Fig. 6). Besides, this function can ensure all the data quality to generate a better virtualization. Duplicating tough checks by humans is inefficient and time consuming. In the worst-case scenario, it is the result of human error.

Furthermore, RStudio is used as well for data cleaning as is an Integrated Development Environment (IDE) that provides a one-stop solution for all statistical computation and graphics. It is a powerful and simple way to engage with R programming. The RStudio is a more advanced version of R that has a multi-pane window setup that allows users to access all the important info on a single screen (such as source, console, environment & history, files, photos, graphs). RStudio was used in this project because various activities

required the use of the R language to help in the visualization and pre-processing of data. The data may be noisy, contain outliers, or simply contain inaccuracies that must be dealt with to increase completeness and compellability. R has excellent data wrangling support. Packages such as dplyr and readr can convert unstructured data into structured data. Therefore, it can improve the efficiency of data pre-processing.

- Handle missing value

```
#=====Identify missing value=====#
#Calculate total missing values
sum(is.na(diabetes))
mean(is.na(diabetes))#Percentage of missing values
colSums(is.na(diabetes))
```

Fig. 7. Handle missing values.

This function identifies the missing value in our dataset (Fig. 7). The sum(is.na()) function shows us how many total missing values there are in this data frame. The mean(is.na()) function displays the overall proportion of missing values in this data collection. Before deal with missing values, must first figure out which columns have missing values. Therefore, may use the colSums(is.na()) function to check.

- Replace missing value with mean

```
#Replacing the missing value with mean
meanAge <- mean(diabetes$Age, na.rm=TRUE)
meanAge
format(round(meanAge, 0), nsma11 = 0)
diabetes$Age[is.na(diabetes$Age)] <-format(round(meanAge, 1), nsma11 = 1)
```

Fig. 8. Replace missing value.

This function will be used when wants to replace the missing value with mean (Fig. 8). It'll start by calculating the mean of the value, then use the format() function to round it to the nearest integer. The is.na() function checks each value to see whether it is NA, and if it is, it replaces the NA values with mean.

- Remove missing value

```
#Remove the missing value
diabetes <- diabetes %>% drop_na(Diabetes_binary)
```

Fig. 9. Remove missing value.

This function will remove any missing values in a specific column (Fig. 9). Pipe operator (%>%) is used to connect the datasets and the select function. The drop\_na() function drops rows that have a missing value in the Diabetes binary columns.

- Remove duplicated data

```
#=====Check duplicate data=====#
duplicated(diabetes)
sum(duplicated(diabetes))
#delete duplicated rows
diabetes <- diabetes %>% distinct()
```

Fig. 10. Remove duplicated data function.

This function removes any duplicated rows from the data frame (Fig. 10). The sum(duplicated ()) function displays the total number of duplicated data in this data frame. The distinct() function can be used to keep just unique or distinct

rows from a data frame. Only the first row is kept if there are duplicate rows.

- Remove outliers of data

```
#-----Identify the OUTLIERS-----#
outvals = boxplot(diabetes$BMI, main = "Boxplot of BMI")$out
#Remove outliers
q1 <- quantile(diabetes$BMI, 0.25)
q3 <- quantile(diabetes$BMI, 0.75)
iqr <- IQR(diabetes$BMI)
diabetes <- subset(diabetes, diabetes$BMI > (q1 - 1.5 * iqr) & diabetes$BMI < (q3 + 1.5 * iqr))
```

Fig. 11. Remove outlier data.

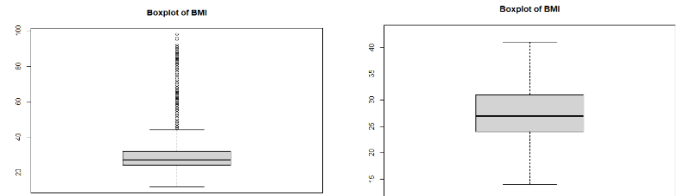


Fig. 12. Boxplot of BMI before and after removing outliers.

Box plots graphically represent the distribution of numerical data and skewness by displaying data quartiles and averages, which can be used to identify outliers within a data collection (Fig. 11 and 12). When analyzing a box plot, an outlier is defined as a data point that lies outside the whiskers of the box plot. The subset() function is then used to remove the BMI values that are more than 1.5 times the interquartile range above the upper quartile and less than 1.5 times the lower quartile (Q1 - 1.5 \* IQR) & (Q3 + 1.5 \* IQR).

- Analysis of data frame's value

```
summary(diabetes)
table(diabetes$Diabetes_012)
```

Fig. 13. Summary function.

```
> table(diabetes$Diabetes_012)
 0      1      2
190055 4629 35097
```

Fig. 14. Table function.

The summary() and table() functions are used to construct and provide the summarized result to the user (Fig. 13 and 14). It is especially handy when the user wants to know the frequency of occurrence of each value in a specific column or data frame. According to the figure above, the table() function is used to analyze the Diabetes 012 column by generating a result in terms of how frequently each value occurs.

```
> summary(diabetes)
Diabetes_012 HighBP Highchol cholCheck BMI Smoker Stroke
Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :12.00 Min. :0.0000 Min. :0.0000
1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:24.00 1st Qu.:0.0000 1st Qu.:0.0000
Median :0.0000 Median :0.0000 Median :0.0000 Median :1.0000 Median :27.00 Median :0.0000 Median :0.0000
Mean :0.2969 Mean :0.429 Mean :0.4241 Mean :0.9027 Mean :28.38 Mean :0.4432 Mean :0.0407
3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:31.00 3rd Qu.:1.0000 3rd Qu.:0.0000
Max. :2.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :98.00 Max. :1.0000 Max. :1.0000
HEARTDISEASERATEK PhysActivity Fruits Veggies myalcoholConsum Anyhealthcare Nodocccost
Min. :0.00000 Min. :0.00000 Min. :0.00000 Min. :0.00000 Min. :0.0000 Min. :0.0000 Min. :0.0000
1st Qu.:0.00000 1st Qu.:1.00000 1st Qu.:1.00000 1st Qu.:1.00000 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:0.00000
Median :0.00000 Median :1.00000 Median :1.00000 Median :1.00000 Median :0.0000 Median :1.00000 Median :0.00000
Mean :0.09419 Mean :0.7585 Mean :0.8343 Mean :0.8114 Mean :0.0562 Mean :0.9511 Mean :0.08418
3rd Qu.:0.00000 3rd Qu.:1.00000 3rd Qu.:1.00000 3rd Qu.:1.00000 3rd Qu.:0.0000 3rd Qu.:1.00000 3rd Qu.:0.00000
Max. :1.00000 Max. :1.00000 Max. :1.00000 Max. :1.00000 Max. :1.0000 Max. :1.00000 Max. :1.00000
```

Fig. 15. Summary function result of diabetes dataset.

The summary() function can be used to generate a summary of the entire data frame by listing out its columns

and displaying the data quartiles, min, max, and average for each variable (Fig. 15).

- Read and Write function

```
#####Read Data#####  
diabetes <- read.csv("C:\\Users\\User\\Downloads\\diabetes_012_health_indicators_BRFSS2015.csv")  
#####Write Data#####  
write.csv(diabetes,"Diabetes.csv",row.names = FALSE)
```

Fig. 16. Read and write function.

These functions are used to read a csv file from the folder that was imported into RStudio and save it to a variable, as well as to export an analyzed data frame from RStudio to a csv file and save it to our computer (Fig. 16). The read.csv function in the code above attempts to read the csv file from the folder and load it into the diabetes variable as a data frame. After completing the data pre-processing, the write.csv function may be used to write the clean diabetes data frame into a csv file called "Diabetes.csv" and exported to a folder.

### C. Data Modeling

Data modelling is the act of developing a visual representation of an entire information system or certain components of it to convey relationships between various data points and organizational structures. The objective is to provide examples of the different types of data that are used and stored inside the system, their relationships, possible groupings and organizational structures, formats, and attributes.

An algorithm is a group of calculations and heuristics used in data mining (also known as machine learning) to build a model from data. The algorithm initially examines the data you submit, searching for kinds of patterns or trends before building a model. In numerous iterations, the algorithm uses the findings of this research to choose the best parameters for building the mining model. The full data collection is then subjected to these criteria to extract useful patterns and thorough statistics. In this section, the algorithm used include decision tree, support vector machine (SVM) and Naïve Bayes algorithm.

1) *Decision tree*: A supervised learning approach that works with both discrete and continuous variables is a decision tree. The dataset is divided into subgroups based on the dataset's most important attribute. The algorithms determine how this attribute is identified by the decision tree and how this splitting is carried out. The root node, which represents the most important predictor, splits into decision nodes, which are sub-nodes, and terminal or leaf nodes, which do not further split.

2) *Support Vector Machine (SVM)*: Finding a hyperplane in an N-dimensional space (N is the number of features) that categorizes the data points clearly is the goal of the support vector machine algorithm (Fig. 17). There are a variety of different hyperplanes that might be used to split the two classes of data points. Finding a plane with the greatest margin, that is, the greatest separation between data points from both classes. Maximizing the margin distance adds some support,

increasing the confidence with which future data points can be categorized.

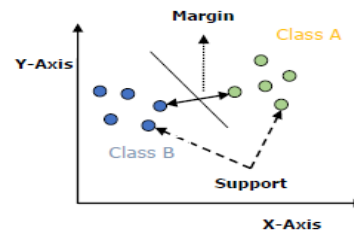


Fig. 17. SVM algorithm.

3) *Naïve Bayes algorithm*: Naïve Bayes algorithm is a classification method built on the Bayes Theorem and predicated on the idea of predictor independence. A Naive Bayes classifier, to put it simply, believes that the presence of one feature in a class has nothing to do with the presence of any other feature. The formula for Bayes' theorem is as follows (Fig. 18):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Fig. 18. Calculation for Bayes' theorem.

- The posterior probability, or P(A|B), measures the likelihood that a given hypothesis (A) will really occur.
- P(B|A) stands for Likelihood Probability, which measures how likely it is based on the evidence at hand that a given hypothesis is correct.
- Priority probability, or P(A), is the likelihood of a theory before seeing the evidence.
- The probability of evidence is marginal probability, or P(B).

## IV. BUSINESS INTELLIGENCE ARCHITECTURE

A business intelligence architecture is the structure that an organization uses to operate business intelligence and analytics applications. It covers the information technology systems and software tools used to gather, combine, store, and analyze BI data before presenting it to high-level executives and other business users as information on daily operations and statistics. The underlying BI architecture is a critical component in implementing an effective business intelligence program that makes data analysis and reporting to assist an organization in tracking business performance, optimizing business processes, identifying new revenue opportunities, improving strategic planning, and providing better decisions overall [4].

Putting such a framework in place helps the healthcare BI team to operate in a coordinated and organized manner to construct an organizational BI solution that fulfils the data analytics requirements of its company. The BI architecture also assists BI and data managers in developing an effective method for processing and handling data that is delivered into the environment (Fig. 19).

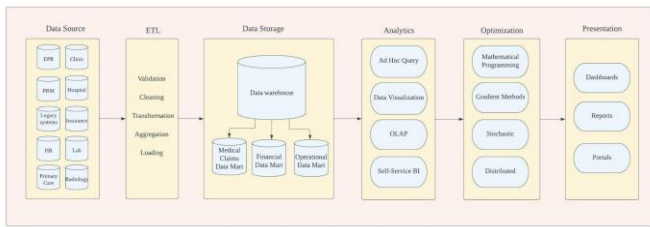


Fig. 19. Proposed BI architecture.

- Data source

These are all the sources that gather and store the data specified as important for the enterprise BI program such as EPR, PRM, radiology, insurance and so on. Secondary sources, such as patient databases from third-party information providers, might also be included. As a result, both internal and external data sources are frequently included in BI architectures. Data relevance, data validity, data quality, and the amount of information in the accessible data sets are all important factors in the data source selection process. Furthermore, to fulfil the data analysis and decision-making requirements of executives and other business users, a combination of structured, semi - structured and unstructured data types may be necessary.

- ETL (Extract, Transform and Load)

ETL is a data integration process that integrates data from numerous data sources into a single, consistent data repository that is then put into a data warehouse. ETL cleanses and organizes data using a set of business rules to fulfil business intelligence objectives, such as monthly reporting, but it may also handle more complex analytics to enhance back-end operations or end user experiences.

- Data Storage

This contains all the repositories where BI data is stored and handled. The most common is a data warehouse that holds structured data in a relational or multidimensional database and allows easy access for querying and analysis. Data warehouse can be linked to smaller data marts which are created for departments data customized to their BI requirements.

- Analytics

In the step, the focus will be on data analysis after handling, processing, and cleaning the data in the previous steps with the aid of a data warehouse. The analytics layer is a series of steps that comprise a toolbox that can be used for any form of analytics. BI application tools are used to meet the pervasive demand for successful analysis to enable organizations of all sizes to develop and earn profit. The four big data analytics techniques include classification, prediction, clustering, and association rules. Besides, a set of technologies may be implemented into a BI architecture to evaluate data and deliver information to business users such as ad hoc query, OLAP and data mining, BI. Specifically in the case of ad hoc query analysis, which allows for higher freedom, flexibility as well as usability in conducting analysis and assisting in the rapid and correct response to crucial business

problems. Moreover, the increased use of self-service BI tools allows managers and business analysts to execute queries on their own rather than depending on members of the BI team to do so. In addition, data visualization tools are also included in BI software and can be used to produce graphical representations of data such like graphs, charts, diagrams and so on to show patterns, outlier elements, and trends in datasets.

- Optimization

In general, an optimization process refers to any process that systematically proposes a better solution and results than previously used solutions. It is the practice of fine-tuning a process to optimize a collection of parameters while keeping within a series of constraints. The main purpose of process optimization is to provide more options for modifying the analytics layer's findings. The optimization block comprises a variety of approaches ranging from mathematical programming to gradient methods and stochastic to distributed.

- Presentation

This layer contains tools that present information to different users in a variety of forms. In the presentation layer, the type of technology includes dashboards, reports, and portals. All these information delivery tools allow business users to see the results and insights of BI and analytics applications for additional data analysis through built-in data visualization and the usual self-service capabilities. Executives and managers who want a broad picture of their organization's performance might use data visualization tools such like dashboards. A dashboard is a handy tool that lets users view data using graphs or charts, colored metrics as well as tables. Furthermore, users may also enable users to visualize more specific information regarding key performance indicators (KPIs) in their organizations. With the help of dashboards, they can more effectively track their progress towards setting goals. Besides, web portals refer to software that makes surfing the internet easier. By using the proposed BI architecture, they can retrieve files in file systems belonging to the Healthcare company or information given by web servers in private networks. Dashboards and web portals may both be configured to enable real-time data access with flexible views and drill-down capabilities. Reports have a more static framework for presenting data.

## V. DATAWAREHOUSE AND OLAP MODEL

The dimensional model above determined the simplest type of data warehouse schema, which is the star schema. This schema is commonly used to design or construct a data warehouse and dimensional data marts. In a data warehouse, a star schema can include one fact table and a series of interconnected dimension tables in the center. In the star schema above, the fact table includes keys to the five-dimension table such as Patient\_ID, Living\_ID, Disease\_ID, HealthCheck\_ID, Service\_ID and other attributes like BMI (Body Mass Index) and diagnostic status. The following are the list of attributes in each of the dimension table:

- Patient dimension table: Sex, Age, Education and Income
- Health service dimension table: AnyHealthcare, NoDocbcCost, CholCheck
- Health check dimension table: High BP, HighChol, GenHlth, MentalHlth, PhysHlth, DiffWalk
- Disease dimension table: Stroke, HeartDiseaseOrAttack
- Living habit dimension table: Smoking, PhysActivity, Fruits, Veggies, HeavyAlcoholConsump.

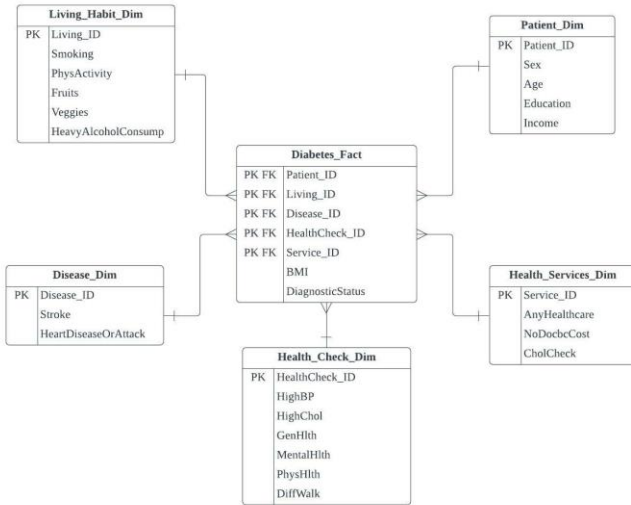


Fig. 20. Star schema.

The advantage of star schema is that all OLAP systems implement the star schema to efficiently create OLAP cubes (Fig. 20). In contrast, many OLAP systems have a ROLAP mode of operation that allows users to use a star schema as a source without having to create a cube structure. Moreover, star schema gains in query performance when compared to fully normalized schemas as it can give improvements in performance for read-only reporting applications.

### VI. DASHBOARD VISUALIZATION

There are two dashboards created for this project which are the diabetes analytics dashboard and patient analysis dashboard. By providing it with a visual context via maps or graphs, data visualization helps us understand what the information means. As a result, it is simpler to spot trends, patterns, and outliers in enormous data sets since the data is easier for the human mind to understand.

#### A. Dashboard for Diabetes Analysis

A dashboard refers to an electronic tracking tool that organizations use to present and summarize their data collection. The dashboard in Fig. 21 shows the information about the overall diabetes analytics according to the Centers for Disease Control and Prevention (CDC). CDC may use a dashboard to determine the rates of diagnosed diabetes, prediabetes, and no diabetes in the general population as well as the leading causes of diabetes. For example, it displays the

number of patients’ diagnostic statuses based on their status of smoking, stroke, heart disease or attack, blood pressure and cholesterol as well as physical activity. Moreover, the dashboard enables CDC to keep track of their patient’s health conditions, so that they can predict the risk and early signs of diabetes and provide the appropriate treatment.

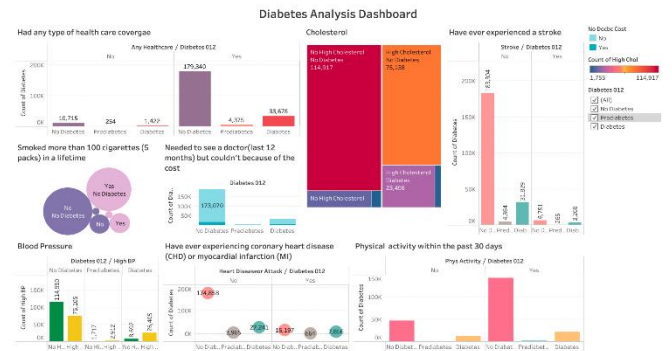


Fig. 21. Diabetes analysis dashboard.

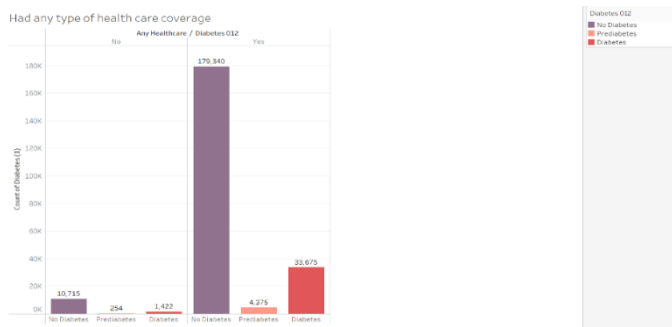


Fig. 22. Healthcare coverage on patient’s diagnostic status.

According to the Fig. 22, a side-by-side bar graph has been used to categorize and present data that results from classifying a group of things based on two or more factors. It shows the number of patients who have or have not had health care coverage based on the patient's diagnostic status (no diabetes, prediabetes, diabetes). The bar graph depicts the patient's diagnostic status using three different colors. Patients without diabetes are represented by purple, those with prediabetes by light pink, and those with diabetes by red. Based on the graph, the number of patients without diabetes with healthcare coverage is 179,340, which is dramatically more than the number of patients without diabetes with healthcare coverage, which is 10,715. As a result, can conclude that patients who have healthcare coverage, such as health insurance or prepaid plans like HMO, can help with diabetes prevention.

The stacked bar graph in Fig. 23 is used to show the count of the patients’ diagnostic status (no diabetes, prediabetes, diabetes) based on the NoDocbcCost in the last 12 months. A stacked bar graph is a type of graph which is utilized to divide and compare parts of a whole. Each bar in the graph represents a whole, and each section represents a different part or category of that whole. The various categories in the bar are represented by different colors. Light blue for no (the patient could see the doctor without worrying about the cost) and dark blue for yes (the patient could not see a doctor because of the

cost). Based on the graph, the number of patients without diabetes who could see a doctor without being concerned about the cost is 173,070, which is noticeably higher than the number of patients without diabetes who could not see a doctor due to the cost, which is 16,985. In addition to that the number of diabetic patients who could see a doctor without worrying about the cost is 31,355, which is higher than the number of diabetic patients who could not see a doctor because of the cost, which is 3,742. As a result, can conclude that most patients do not have to be worried about the cost of seeing a doctor, and that it has little impact on a patient's risk of developing diabetes.

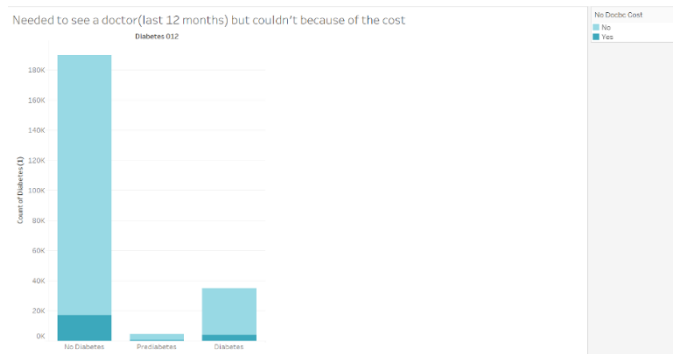


Fig. 23. Count of patients' diagnostic status based on medical fees.

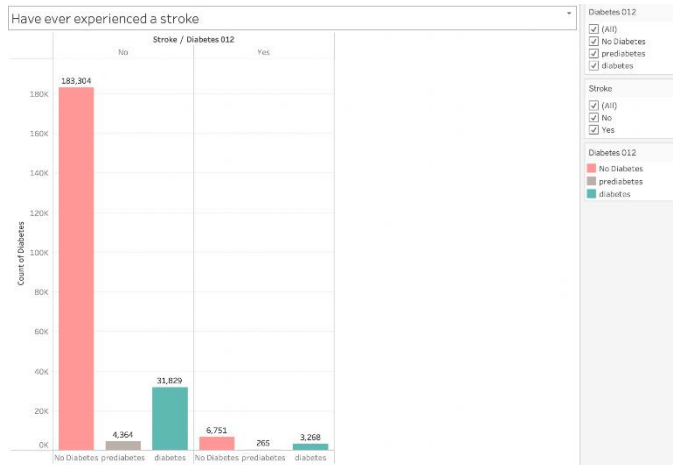


Fig. 24. Patients' diagnostic status based on stroke.

Fig. 24 represents the patients' diagnostic status which includes no diabetes, prediabetes, diabetes based on stroke and non-stroke. There are two categorical variables which are displayed using the side-by-side column graph. A side-by-side column graph is used to illustrate two category variables. The results collected when the patients are classified according to diagnostic status (no diabetes, prediabetes, diabetes) and the state of stroke (has stroke, no stroke) are displayed in the side-by-side column graph above. The bar graph uses three unique colors to indicate the patient's diagnostic status which is categorized by the diabetes status. Pink color for no diabetes, grey color for prediabetes and green color for diabetes. The filter option in the upper right corner can help you select and display only the data you want to view, making comparison and analysis easier as well. Furthermore, also observe that no diabetes has the highest count within no diabetes which is

183304 out of 219497, and no diabetes has the highest count within stroke which is 6751 out of 10284. Besides that, can see that the number of patients with diabetes and prediabetes who have no stroke is higher than the number of patients who have stroke, reaching 36193 respectively, compared to the number of patients who have heart disease and attack, which is 3798. To conclude, stroke is not the main reason to get diabetes because the number of patients with non-stroke is higher than the patients with stroke.

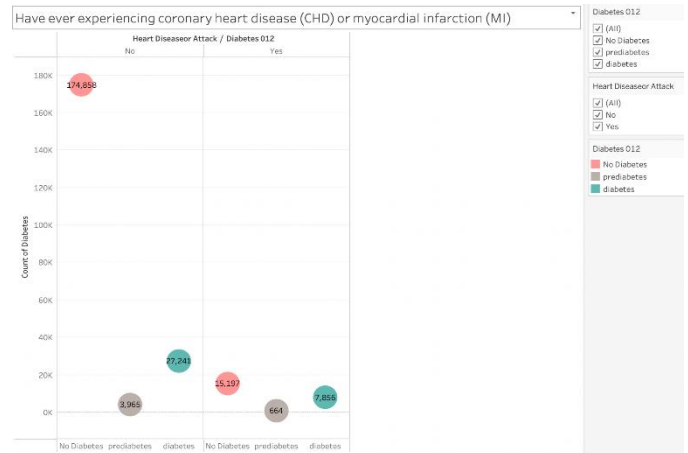


Fig. 25. Patients' diagnostic status based on heart disease.

Fig. 25 represents the patients' diagnostic status which includes no diabetes, prediabetes, and diabetes, based on whether the patient has heart disease and attack, and non-heart disease and attack. A Multi-Category Chart is useful when you have data for components that fall into multiple categories. With this, can show comparative data in more than one category. You can distinguish various states based on their varied colors using the card in the upper right corner. Based on the graph above, pink represents no diabetes, grey represents prediabetes, and green represents diabetes. It shows that the highest count inside no diabetes is 174 858 out of 206064, while the greatest count within stroke is 15197 out of 23717. Patients who get heart disease and attack with no diabetes accounted half of the total. However, can observe that the count of diabetes and prediabetes in no heart disease or attack is higher than the patients who had heart disease or attack, reaching 31377 respectively, compared to the count of those who have heart disease and attack, which is 8520. From this, it can be stated that heart disease rarely affects diabetes.

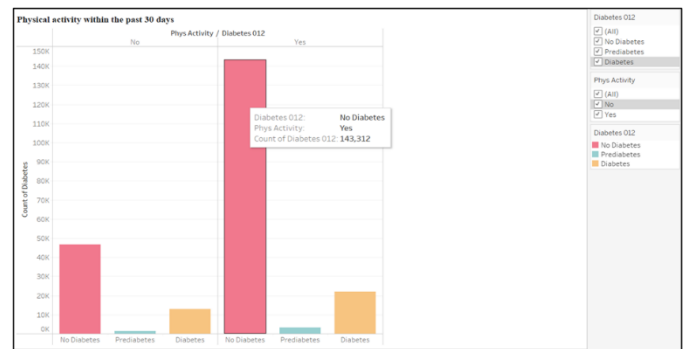


Fig. 26. Patients' diagnostic status based on physical activity.



Fig. 26 depicted the count of patients' diagnostic status (no diabetes, prediabetes, diabetes) based on physical activity and no physical activity. The two categorical variables are displayed using side-by-side column graph. A side-by-side column graph can be used to organize and present data that results from categorizing a group of people or things using two or more criteria. For example, the side-by-side column graph above displays the data obtained when patients are categorized according to diagnostic status (no diabetes, prediabetes, diabetes) and state of exercise (has physical activity, no physical activity). The bar graph uses three unique colors to indicate the patient's diagnostic status. According to the legend, the pink column represents patients who do not have diabetes, the green represents patient who have prediabetes, and yellow represent patients who have diabetes. The filter option in the right corner can assist in filtering and displaying only the data that you want to see, making comparison and analysis easier. Based on the graph, it demonstrates that the number of patients without diabetes who engage in physical activity is 143312, which is significantly higher than the number of patients without diabetes who do not engage in physical activity, which is 46743. From this, it can be assumed that physical activity can aid in the development of resistance and the prevention of diabetes.

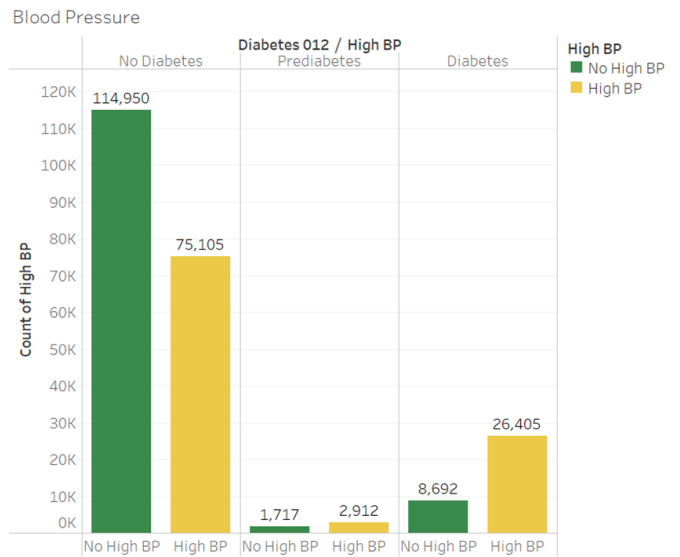


Fig. 28. High blood pressure status based on patient's diagnostic status.

Fig. 28 shows the count of high blood pressure and no high blood pressure based on the patient's diagnostic status (no diabetes, prediabetes, diabetes). A side-by-side bar graph is used to organize and present data that results from categorizing a group of things according to two or more criteria. The bar graph displays two distinct colors to depict whether the patient has blood pressure. Patients with no high blood pressure are represented by green, while those with high blood pressure are represented by yellow. From this graph, the number of patients without diabetes and having high blood pressure is 75105 which is much lower than the number of patients without diabetes and no high blood pressure at 114950. However, it depicts that for prediabetic and diabetic patients, the count of high blood pressure is relatively high, reaching 2912 and 26405 respectively, compared to the count of no high blood pressure, which is 1717 and 8692.

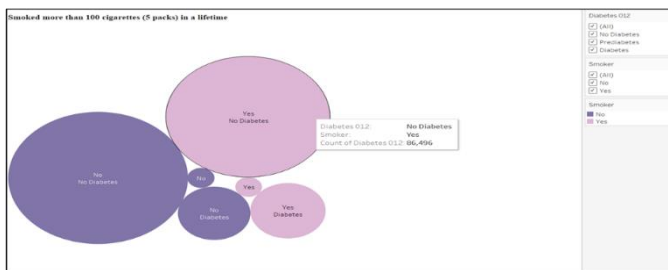


Fig. 27. Association between the diagnostic status and smoking.

In Fig. 27 above, a cluster of circles represents the association between the diagnostic status (no diabetes, prediabetes, diabetes) and smoking. To make the best use of space, the bubbles are packed in as tightly as possible. The individual bubbles are defined by the category field, and the bubble size is represented by the value field. Above, it shows at diagnostic status and smoking. While, can't control how the bubbles are arranged, but can control how big they are by putting a measure on size, in this case, and use the count of diagnostic status. The size of the bubbles represents the count of diagnostic status for various combinations of diagnostic and smoking status. This function is especially beneficial since the size of the bubble clearly shows the difference between them. The larger the circle, the larger is the proportion of the total. When a legend field is assigned to the Packed Bubble chart, the grouping mode is selected by default. It colorizes and divides the bubbles further based on the legend field. The color scheme is classified into two categories: smokers and non-smokers. According to the graph, most diabetic patients are heavy smokers (18223), while the majority of non-diabetic patients are non-smokers. Therefore, may conclude that smoking has a significant impact on causing diabetes.

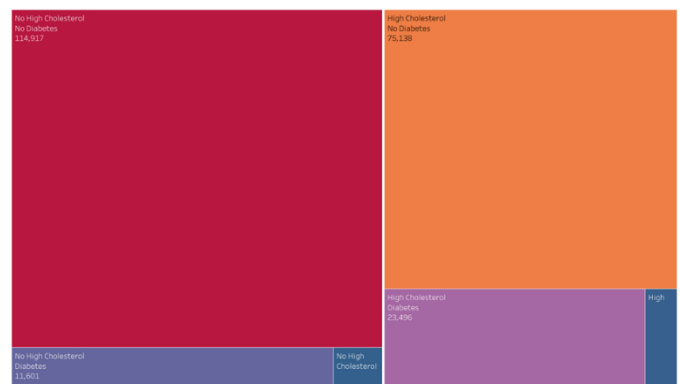


Fig. 29. High cholesterol status based on patient's diagnostic status.

As depicted in Fig. 29 above, it shows high cholesterol and non-cholesterol counts according to the patient's diagnostic status. The tree map is a visual representation consisting of nested rectangles. These rectangles indicate different categories within a given dimension and are arranged in a tree-like hierarchy. The tree map chart displays various distinct colors to distinguish between high cholesterol and non-

cholesterol in different diagnostic statuses. When looking for insights in a tree map, the largest box represents the largest portion of the entire, while the smallest box represents the smallest portion.

Fig. 30 below depicts the total number of no high cholesterol and high cholesterol for each diagnostic status such as no diabetes, prediabetes, and diabetes. From this table, we can analyze whether high cholesterol has a strong effect on causing diabetes.

High Chol	No Diab..	Prediab..	Diabetes
No High Cholesterol	114,917	1,755	11,601
High Cholesterol	75,138	2,874	23,496

Fig. 30. Total number of high cholesterol status for each diagnostic status.

### B. Dashboard for Patient Analysis

The spreadsheets and dashboards shown below are mainly focused on patient-related data, starting from the patient's gender, age group, education level and their income. The purpose is to obtain a broad perspective of patient's personal information and acquire relevant knowledge for Centers for Disease Control and Prevention (CDC) management to analyze whether these personal factors will cause diabetes and may use that as a reference to make a better decision for prevention.

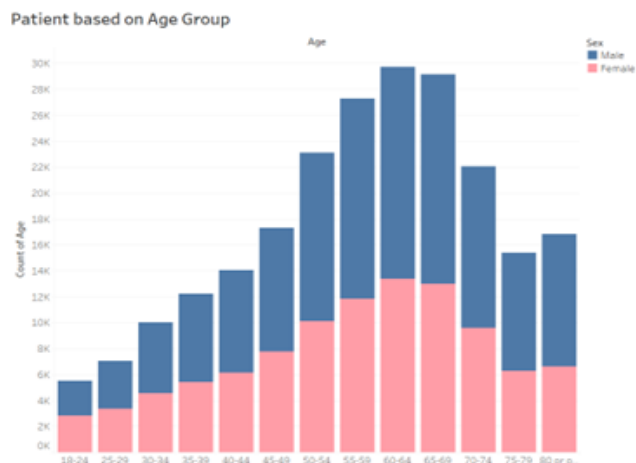


Fig. 31. Patient based on age group.

In Fig. 31 above represented the horizontal bar of patient based on their age group separated by gender. The stacked bar graph can differentiate the gender based on their distinct colors, which is described at the card on the top right corner. If click one of the genders in the horizontal bar chart, it will direct us to another worksheet which is specifically filter out based on the gender have chosen by using filter action.

The figures are the age group filtered out by gender (Fig. 32). The left figure represents the age group by male while the right figure represents the age group by female. Tableau action filters enable us to transfer information between worksheets (Fig. 33). Generally, when you choose marks from one worksheet, that information is sent to other worksheets, resulting in the display of the relevant information. Behind the scenes, action filters deliver data

values from appropriate source fields to the destination worksheet as filters. By implementing this function, the CDC management will be able to identify and focus the age groups based on gender.

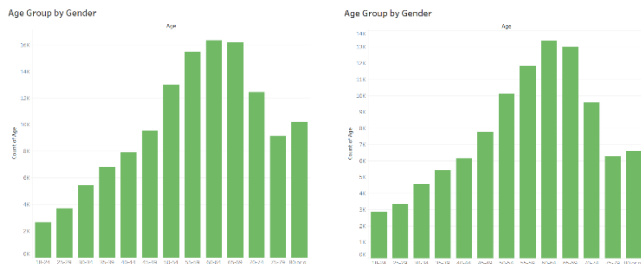


Fig. 32. Total number of high cholesterol status for each diagnostic status.

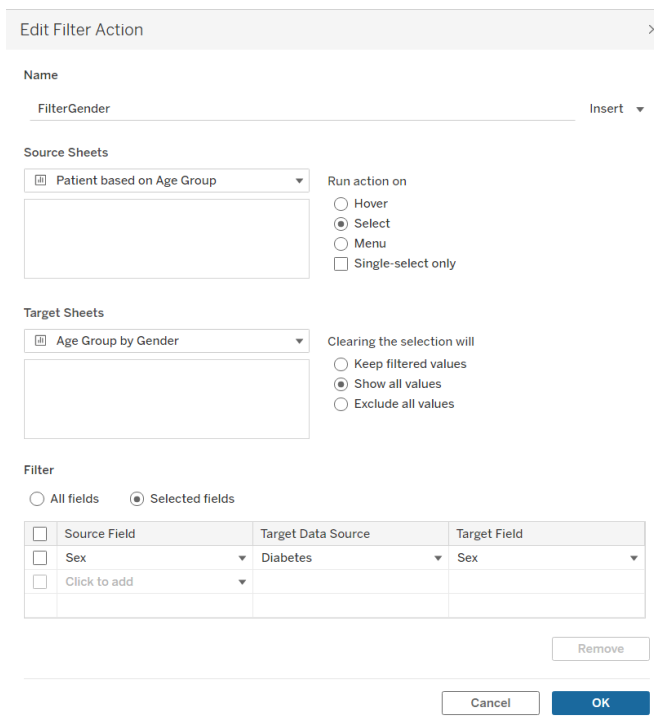


Fig. 33. Filter action in tableau.

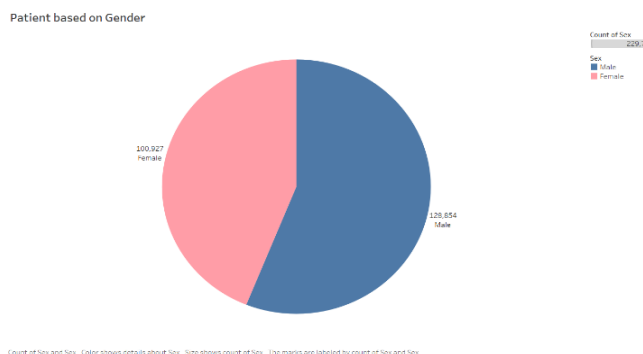


Fig. 34. Patient based on gender.

The pie chart above (Fig. 34) shows the number of patients based on gender. The pie chart displays two different colors, each representing a gender. Male is represented by blue, while female is represented by pink. This provides a general

estimate of the proportion of male and female customer patient collected from the database. The figure can clearly find that male has a greater number than female, which male has several 128,854 and 100,927 for females.

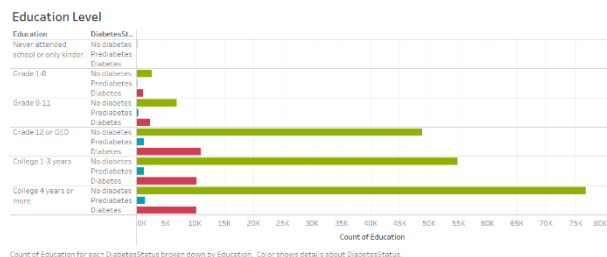


Fig. 35. Education level.

Fig. 35 above determines the highest education level patient had taken. There are six level of education which are never attended school or only kindergarten, Grade 1 through 8, Grade 9 through 11, Grade 12 or General Education Diploma (GED), College 1 year to 3 years, College 4 years or more. It basically shows whether the level of education play a role in diabetes. For instance, the number of patients whose highest education level is 4 years college or more and does not diagnose with diabetes are 76,746 people. Studies shows that education levels may increase the adoption of health behaviors such as sufficient eating and medication adherence. As a result, it is likely that education levels work as a fundamental cause of disease by using resources such as knowledge, which have a substantial effect on people's ability to decrease risks that may prevent or delay diabetes or better treat the disease once it comes.

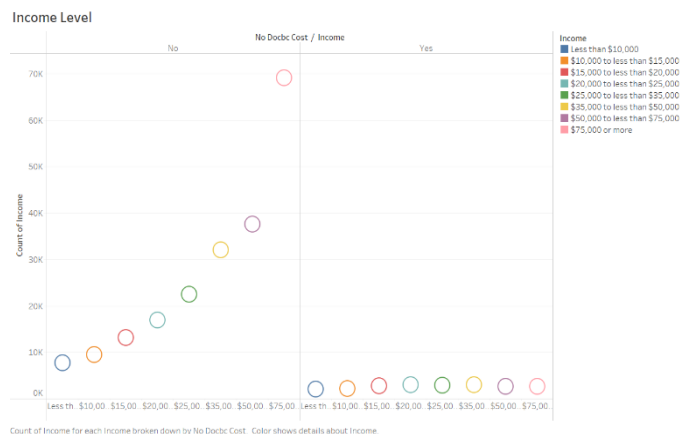


Fig. 36. Income level.

Fig. 36 above determines the whether the patient has a time in the past 12 months when needed to see a doctor but could not be due to the cost of treatment based on their income level. Studies shows that people who have low income likely to have a higher risk of diagnose with diabetes. It can find there is a big different between the side-by-side circle visualization. As a result, this can be assumed that most patients who have an income of \$75,000 or more do not have to worry about the treatment cost, that is, 69,117 people (see Fig. 37).

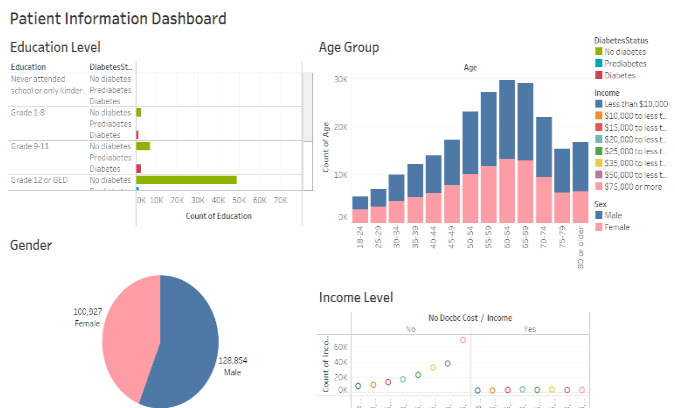


Fig. 37. Patient analysis dashboard.

## VII. CONCLUSION

The goal of Business Intelligence is to assist and enable better business decisions. BI provides organizations with access to information that is crucial to the success of a variety of areas and departments. Effectively integrating BI will give the organization with more actionable data, valuable insights into market trends, and a more strategically orientated decision-making approach. In this project, the author has investigated Centers for Disease Control (CDC) to create a BI solution for them to make better decision after visualization the insight from the dataset.

Before visualizing the data, author must do some in-depth research on the domain background, their problem, and opportunities, create a project charter, data identification, designing BI logical architecture as well as data warehouse model. Once all of this has been done, the author will start building dashboards that can develop insights to the executive and assist them in making a better decision by using Tableau. Tableau is a powerful tool for quickly creating interactive data visualizations and user-friendly. The author has created 2 dashboards which contain patient information analysis and health information analysis.

Finally, BI solution is important as it aids in the production of reliable reports by collecting data directly from the data source. Today's BI solutions minimize the time-consuming effort of manually aggregating data since BI technologies provide up-to-date data, executives can monitor firms in real time. As a result, the author thinks that the project's results might be improved further for developing a best-fit BI solution and data visualization for CDC. With the enhancement and improvement, it will undoubtedly assist CDC in striving for greater opportunities in terms of patients, health and so on.

## REFERENCES

- [1] "Benefits of Switching to an Electronic Health Record (EHR)," *Practice Fusion*, Nov. 22, 2016. <https://www.practicefusion.com/health-informatics-practical-guide-page-1> (accessed Sep. 21, 2022).
- [2] "4 Health Care Data Challenges and How to Overcome Them," *Corporate Compliance Insights*, Jun. 28, 2018. <https://www.corporatecomplianceinsights.com/4-health-care-data-challenges-overcome> (accessed Sep. 21, 2022).

- [3] “NHANES - About the National Health and Nutrition Examination Survey,” *Centers for Disease Control and Prevention*, 2019. [https://www.cdc.gov/nchs/nhanes/about\\_nhanes.htm](https://www.cdc.gov/nchs/nhanes/about_nhanes.htm)
- [4] “What is Business Intelligence Architecture (BI Architecture)?,” *SearchBusinessAnalytics*. <https://www.techtarget.com/searchbusinessanalytics/definition/business-intelligence-architecture>
- [5] A. Menon, “What is Agile Business Intelligence? - A Guide,” *Techfunnel*, Aug. 08, 2021. <https://www.techfunnel.com/information-technology/agile-business-intelligence> (accessed Sep. 21, 2022).
- [6] JavaTpoint, “Agile Model - javatpoint,” *www.javatpoint.com*, 2011. <https://www.javatpoint.com/software-engineering-agile-model>
- [7] Wikipedia Contributors, “Centers for Disease Control and Prevention,” *Wikipedia*, Dec. 09, 2018. [https://en.wikipedia.org/wiki/Centers\\_for\\_Disease\\_Control\\_and\\_Prevention](https://en.wikipedia.org/wiki/Centers_for_Disease_Control_and_Prevention)
- [8] “Business Intelligence in Healthcare,” *Villanovau.com*, 2022. <https://www.villanovau.com/resources/bi/business-intelligence-in-healthcare/#:~:text=By%20providing%20a%20foundation%20for> (accessed Sep. 21, 2022).
- [9] HealthIT.gov, “What are the advantages of electronic health records?,” *HealthIT.gov*, Mar. 08, 2022. <https://www.healthit.gov/faq/what-are-advantages-electronic-health-records>
- [10] D. Dugar, “Benefits of Electronic Health Records | EHR Advantages & Disadvantages,” Sep. 19, 2021. <https://www.selecthub.com/medical-software/benefits-of-ehr-systems> (accessed Sep. 21, 2022).
- [11] “What you need to know about BI dashboards,” *Tableau Software*. <https://www.tableau.com/learn/articles/business-intelligence/bi-dashboards>
- [12] University of Massachusetts Dartmouth, “Decision-making process - UMass Dartmouth,” *Umassd.edu*, 2021. <https://www.umassd.edu/fycm/decision-making/process/>
- [13] A. B. I. BV, “Where do Clinical Evaluation and Clinical Investigation meet?,” *www.qservegroup.com*. <https://www.qservegroup.com/eu/de/i319/where-do-clinical-evaluation-and-clinical-investigation-meet>
- [14] Heather, “7 Major Challenges Facing the Healthcare Industry in 2021,” *MailMyStatements*, Oct. 27, 2020. <https://mailmystatements.com/2020/10/27/2019challenges/>
- [15] R. College, “4 Major Challenges Facing Today’s Health Care Industry,” *Regis College Online*, Mar. 10, 2018. <https://online.regiscollege.edu/blog/4-challenges-facing-the-health-care-industry/>

# Augmented, Virtual and Mixed Reality Research in Cultural Heritage: A Bibliometric Study

Nilam Upasani<sup>1</sup>, Asmita Manna<sup>2</sup>, Manjiri Ranjanikar<sup>3</sup>

Balaji Institute of Technology and Management, Sri Balaji University, Pune, India<sup>1</sup>

Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India<sup>2,3</sup>

**Abstract**—Heritage allows us to learn about the different monuments of importance and the inherited traditions from our ancestors. However, many times the monuments get partly ruined due to natural wear and tear, sometimes due to attacks by invaders. To preserve the cultural heritage virtually, many researchers have used augmented, virtual and mixed reality for bringing ancient environments to live in those heritage sites. This study aims to identify publications related to virtual, augmented and mixed reality in cultural heritage and to present the bibliometric analysis of these studies. The research articles on virtual, augmented, and mixed reality in cultural heritage are retrieved using the Scopus database. The analysis is performed using VOSviewer and various parameters such as bibliographic coupling of the countries, publications, journals, authors, and co-occurrences of the author keywords are performed. From the analysis done in this study, it is discovered that the augmented, virtual and mixed reality research in the domain of cultural heritage is mostly concentrated in Italy and surrounding European countries. However, it is also found that the research in this domain is lagging in many countries even if those countries are the homes of various heritage sites. This study provides an extensive analysis of the recent literature related to augmented, virtual and mixed reality research in cultural heritage. This information science based analysis will help researchers to identify the prominent journals in this domain, recognize stalwarts in this field and follow their works, find out path-breaking publications to refer to, and predict the direction of future studies.

**Keywords**—Augmented reality; bibliometric analysis; cultural heritage; information science; mixed reality; virtual reality

## I. INTRODUCTION

For the essential growth of a community, its history plays a vital role. Heritage allows us to learn about the history, culture and inherited traditions from ancestors by observing the monuments, archaeological sites, forts, historical city centres, museums, etc. [1][2][3]. Millions of tourists visit the historical sites every year to get a feel of individual and collective identity. It contributes to the cohesion of various cultures across the world. This feel of cohesion can be enhanced by combining the real and digital content to provide sensory experiences. Depending upon the involvement of the user in this virtual environment, the technologies can be classified into three types namely, immersive, non-immersive and semi-immersive.

Immersive technologies create a virtual world by combining the physical world with a digital or simulated reality such that the user feels that (s)he is actually engrossed in that

virtual world. On the other hand, non-immersive technologies also generate a virtual world for the user; however, the complete control of that virtual world lies with the user only. Semi-immersive technologies create a partial virtual environment that allows users to feel the virtually created world and remain aware about the physical surroundings simultaneously.

Augmented Reality (AR) is an immersive technology where real-world objects are upgraded with computer-aided non-cognitive visual and/or auditory information. Virtual Reality (VR) is a simulated experience that might be comparable to or wholly different from reality [4]. Mixed Reality (MR) is a combination of physical reality, AR, and VR. Virtual and mixed reality experiences can be fully immersive, semi-immersive or non-immersive. There are many applications of AR, VR and MR that include Education (such as medical or military training), Entertainment (such as movies, and video games), Business (such as virtual meetings), Tourism (virtual tours of cultural heritage and ancient structures sites), etc. [5].

Rejuvenating the demolished structures with augmented, and virtual reality helps the younger generation to understand the rich heritage of ancestral civilizations, their customs and traditions, the culture followed, and the stories of bravery and ethics. Storytelling using virtual reality enriches the learning experience of the young generation for exploring the rich heritage.

With the advent of mixed reality, countries having rich cultural heritage are coming up with mixed reality-based projects as well to preserve and protect their heritage architecture and assets.

### A. Motivation and Contribution

With more researchers taking up similar projects, especially to preserve the cultural heritage of respective countries, more and more young people are getting aware of their own culture and tradition. However, there is a lack of thorough analysis of the difficulties in the field's research or its future directions. This study is done to review the current status of research in cultural heritage connected to AR, VR, and MR.

Research objectives of this paper are to:

1) Compile the peer-reviewed articles and conference papers on applications of AR/VR in cultural heritage during the period of 1999 to 2022.

- 2) Ascertain the spread of this field using bibliometric coupling of publications, journals, and authors.
- 3) Understand which countries have the most publications in this area.
- 4) Which technologies were employed for the implementation of cultural heritage projects.

This is accomplished through the use of bibliometric analysis. As defined by Pritchard, bibliometrics is “the application of mathematical and statistical methods to books and other media of communication” [48]. In this study, bibliometric analysis has been carried out for classifying the publications based on AR/VR in cultural heritage. The bibliographic information, like citations, keywords, link strengths, countries, journals, and authors of a publication has been considered for the analysis. Good quality publications have been selected from the widely-used and well-recognized online Scopus database for this study. This database includes almost all the important research papers in this domain. After retrieving the data, the same has been analyzed using the bibliometric visualization tool called VOSviewer.

### B. Organization of the Paper

Section II presents a summary of cultural heritage projects where AR, VR, or MR has been deployed; Section III discusses the research methodology used for the bibliometric analysis; the results are analyzed in Section IV; Section V discusses the insights obtained from the results; are conclusions are available in Section VI.

## II. SUMMARY OF CULTURAL HERITAGE PROJECTS

This section presents a survey of various immersive, non-immersive, and semi-immersive projects implemented across the globe. The category-wise distribution is shown in Table I.

TABLE I. CATEGORIZATION OF CULTURAL HERITAGE PROJECTS

Project	Reference
Immersive	[6-8], [9-11], [12], [13-14], [15-16], [17], [18], [19-20], [21-22], [23-24]
Non-immersive	[25], [26], [27-28], [29], [30-31], [32-33], [34]
Semi-immersive	[35], [36-40], [41-42], [43], [44]

In the “Museum of Pure Form” [6-8] project, a virtual museum was created where users are allowed to have an immersive experience of interacting with an art piece by nestling before that art piece. A project entitled “CREATE” [9-11] was developed for creation and maneuvering of virtual worlds by using actual data source and integrating auditory and other haptic senses. Another research in Japan [12], included an interaction system with sight, hearing, smell, and touch for a digital museum. The “National Archaeological Museum” of Italy’s Marche offered a novel blend of actual and virtual settings, as well as an immersive solution, to improve the understanding, knowledge, and holistic perception of museum visitors [13-14]. A project called “The Feelies” attempted to create a multi-sensory solution in 2015 using virtual reality in theatre settings, with a focus on developing, filming, and distributing quality content [15-16]. The “Interactive Haptic System for Archery” was designed with the objective of constructing engineering solutions for training purposes

through virtual reality simulations, enabling users to engage in classical archery [17]. The project “Thresholds,” showcased in the UK and Turkey, generated a virtual world that could be researched in a space that was emulated in person keeping with the virtual scenario [18]. The project “Zelige Door on Golborne Road” in London emphasised the connection between interaction design and people’s perceptions by providing an augmented reality app with sensory technologies to overlay pre-recorded video and aromas [19-20]. The “M5SAR” project’s approach for a mobile multi-sensory AR system for museums aimed to create an AR system that could serve as a tour guide for historical and cultural events [21-22]. The project “SensiMAR” proposed an AR based mobile experience outdoors that would render additional information during searching an archeological location in Portugal [23-24].

The “Haptic Museum” was a groundbreaking experiment that provided visitors a non-immersive experience by providing an opportunity to study 3-D works of art by virtually “touching” those art pieces via the internet [25]. In the Civic Museum of Como, Italy, the temporary museum exhibit named “The Fire and the Mountain” was designed with the intention of promoting active and social learning through a variety of media and multi-sensory experience [26]. Commercial haptic devices were incorporated for visitors to observe, hear, and feel virtual equivalences of the actual objects for a virtual exhibition in Colombia [27-28]. A virtual museum called “Museu3I” was created to enable users to conduct virtual tours using haptic devices [29]. A project entitled “Hapto-visual and Auditory Rendering,” combined the senses of sight, hearing, and touch for museum visitors [30-31]. The multisensory study “The Reconstructed Historical City of Tomis,” was developed to provide various assistance with the help of haptic devices in Romania [32-33]. A study on one more project to provide a multisensory experience was presented in [34].

The “SenSpace” initiative, deployed and tried in the USA, employed tactile, auditory, and visual evidence to help users understand the Narcissus story of Greek mythology in a physical environment in a semi-immersive way [35]. The project “MediaEvo” is a gaming application to examine the rebuilt historical setting of the Middle Ages through VR [36-40]. Another semi-immersive project called “Virtual Kyoto” provided users an opportunity to experience the Gion Festival, in Japan, with the help of optic and tactile senses, high-quality festival music, and a virtual Yamahoko float ceremony in 3D [41-42]. Some more studies on the projects that improved user experience statistically are presented in [43-44].

As observed from Table I, most of the museum tour related projects are based on immersive technology and comparatively less work is done in non-immersive and semi-immersive areas. Digging down further, it can be found that these projects mainly rely on stimulating the optic senses of users, followed by the haptic, auditory, smell, and taste senses.

Other than the domain of museology, AR, VR has widely been used in the reconstruction of demolished / semi-demolished historical sites as well. One of the most talked about reconstruction project was the Rome Reborn project where the concept of virtual reality was used to demonstrate the urban development of ancient Rome [45].

To preserve the virtual heritage of the first Afro-American Church of the Indianapolis, USA, the 3D reconstruction project “Virtual Bethel” was taken up by Indiana University, USA [46].

Another ambitious 3D-reconstruction project of an ancient monument in Istanbul, Turkey was the “Virtual Hagia Sophia” developed by researchers from MIRALab at the University of Geneva [47].

From the above discussion, it can be understood that AR, VR and MR have widely been employed in the domains of 3D reconstruction and in providing immersive or semi-immersive experiences to tourists at different historical sites and museums. However, there is a dearth of comprehensive studies for ascertaining the spread of this field in terms of good journals, pathbreaking researchers, technologies used by various researchers, and so on and so forth. This study is aiming to find such answers.

### III. RESEARCH METHODOLOGY

The various databases available for retrieving the research work available in different domains are Scopus, ACM Digital Library, ERIC, IEEE Xplore, PubMed, ScienceDirect, Web of Science etc. Out of these databases, ERIC is specific to research articles related to Education, PubMed hosts papers related to medical science, IEEE Xplore hosts research articles related to engineering, while ACM Digital Library focuses primarily on Computer Science related articles. Scopus, ScienceDirect and Web of Science are multidisciplinary databases. For this bibliometric analysis, the Scopus database has been chosen as this is a widely accepted and reliable database. The database was searched on 11<sup>th</sup> August 2022 to find publications related to AR/VR in cultural heritage and 3D reconstruction.

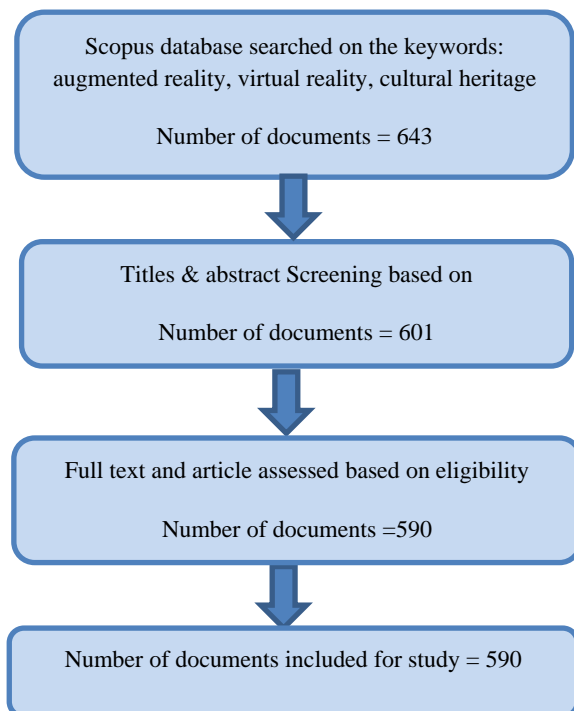


Fig. 1. PRISMA diagram.

Research articles in the duration of 1999 to 2022 have been considered for this study. Based on this search criteria, a total of 643 publications are acquired from the Scopus database. The author, affiliation, journal, year of publication, keywords, and counts of citations are all exported into CSV format for these 643 papers. However, a manual review is needed to assess whether all the retrieved publications are pertinent to the topic. The detailed process to finalize the list of documents as a part of the systematic review [49], is shown as a “Preferred Reporting Items for Systematic Reviews and Meta-Analyses” (PRISMA) flow diagram in Fig. 1.

As shown in Fig. 1, after scrutinizing the titles and abstracts of the papers, some of the publications related to 3D reconstructions in the medical surgery domain are removed resulting in 601 documents. Out of 601 documents, some are removed after full text analysis based on various parameters such as not sufficient data, no diagnostic criteria, etc. So, the final bibliometric analysis is presented based on 590 publications. The year-wise distribution of these documents is shown in Fig. 2.

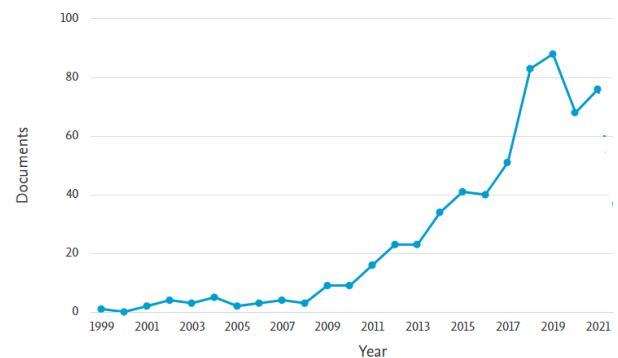


Fig. 2. Year-wise distribution of the shortlisted documents.

From Fig. 2, it can be observed that even if the research started before the year 2000, it started getting more attention from the year 2015 and it gained actual momentum from 2018 onwards.

As per the study, VOSviewer and CitNetExplorer are among the most popular computer visualization tools for bibliometric analysis [50-53]. Here, we have used VOSViewer version 1.6.17 for constructing and visualizing the bibliometric networks for analyzing the bibliographic information, like citations, keywords, link strengths, countries, journals, and authors of a publication.

### IV. RESULTS

This segment discusses the results of various bibliometric analyses. The term ‘Bibliographic coupling’ was coined by Kessler in the early 1960s [54-56]. It is a method for grouping scientific and technical publications. In this paper, the Bibliographic coupling of countries, journals, authors, publications, and author keyword co-occurrences of identified publications, discussed in Section 2, are analyzed. The results are presented in such a manner that helps researchers to understand the linkage by starting with more general information like major countries contributing research in this

domain to more specific information such as the most cited author or keywords-coupling of different publications.

A. Bibliographic Coupling of the Countries

A bibliographic connection of the countries is shown in Fig. 3 and Table II, along with a network visualization. There is a ten-country restriction per document. A country's minimum requirement for publications is 12. Thirteen of the seventy-one countries met the requirements. For each country, the number of documents, citations, and total link strength is computed. The countries with the most total connections are selected. Italy came in first with 157 publications, 1489 citations, and 3398 overall link strength. Rest countries are; United Kingdom (53; 1546; 2533), Spain (45; 338; 2121), Greece (61; 733; 2027), Portugal (29; 151; 1757), France (27; 354; 1124), Australia (17;

348; 953), South Korea (25; 352; 917), Germany (28; 296; 722), China (30; 61; 641), Malaysia (12; 48; 601), United States (22; 120; 498), and Indonesia (13; 39; 154). The first figure inside the parentheses indicates the number of documents, the second indicates the total citations, and the total strength of the other nations' links is indicated by the third figure.

Different colours in Fig. 3 represent different clusters that are more commonly related to one another. It indicates that studies from nations in the same cluster are more likely to cite each other. Spain, Portugal, France, Germany, Greece, UK, and China, make up the largest cluster. The second cluster includes Australia, Indonesia, Italy, and Malaysia. The third cluster includes South Korea and the United States.

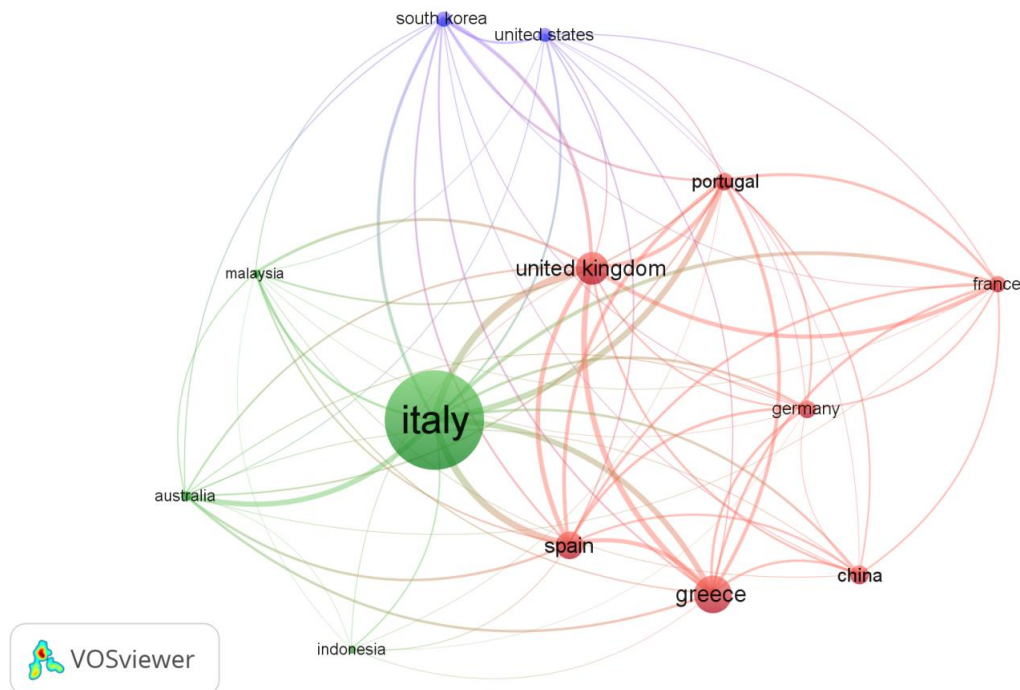


Fig. 3. Countries bibliographic coupling (network visualization).

TABLE II. COUNTRIES BIBLIOGRAPHIC COUPLING

Country	Documents	Citations	Total link strength
Italy	157	1489	3398
United Kingdom	53	1546	2533
Spain	45	338	2121
Greece	61	733	2027
Portugal	29	151	1757
France	27	354	1124
Australia	17	348	953
South Korea	25	352	917
Germany	28	296	722
China	30	61	641
Malaysia	12	48	601
United States	22	120	498
Indonesia	13	39	154



### B. Bibliographic Coupling of the Journals

The bibliographic connection of the journals with density visualization is depicted in Fig. 4. As an inclusion criterion, the journal should have at least a specific number of articles. Only ten journals out of 288 passed the test. For those ten journals, the number of articles, citations, and total strength of bibliographic coupling relationships with other journals are calculated. The journals with the most total link strength are chosen and arranged as per the total link strength as shown in Table III.

Each circle in Fig. 4 depicts a journal, with varying colors representing the density of the journal. The density visualization is weighted based on the number of articles for each journal. The colors shifting to yellow and then red indicate that the associated journal has published more articles.

With 94 publications, 711 citations, and 647 total link strengths, the Lecture Notes in Computer Science series is the

most popular. The “Journal on Computing and Cultural Heritage” is ranked second in terms of citations (396), overall link strength (429), but fifth in terms of content count (13 publications).

The second-highest number of articles (26), with the third-highest citations (202) and a link strength of 137, are published in “International archives of photogrammetry, remote sensing, and spatial information sciences”.

### C. Bibliographic Coupling of the Authors

Fig. 5 uses overlay visualization to demonstrate the authors' bibliographic connection. An author needs to have a minimum of six publications to be considered for this study. Only ten authors are chosen from a total of 1672. For each author, the number of documents, citations, and total link strength is calculated. The authors with the most total link strengths are chosen and arranged as per the total link strength as shown in Table IV.

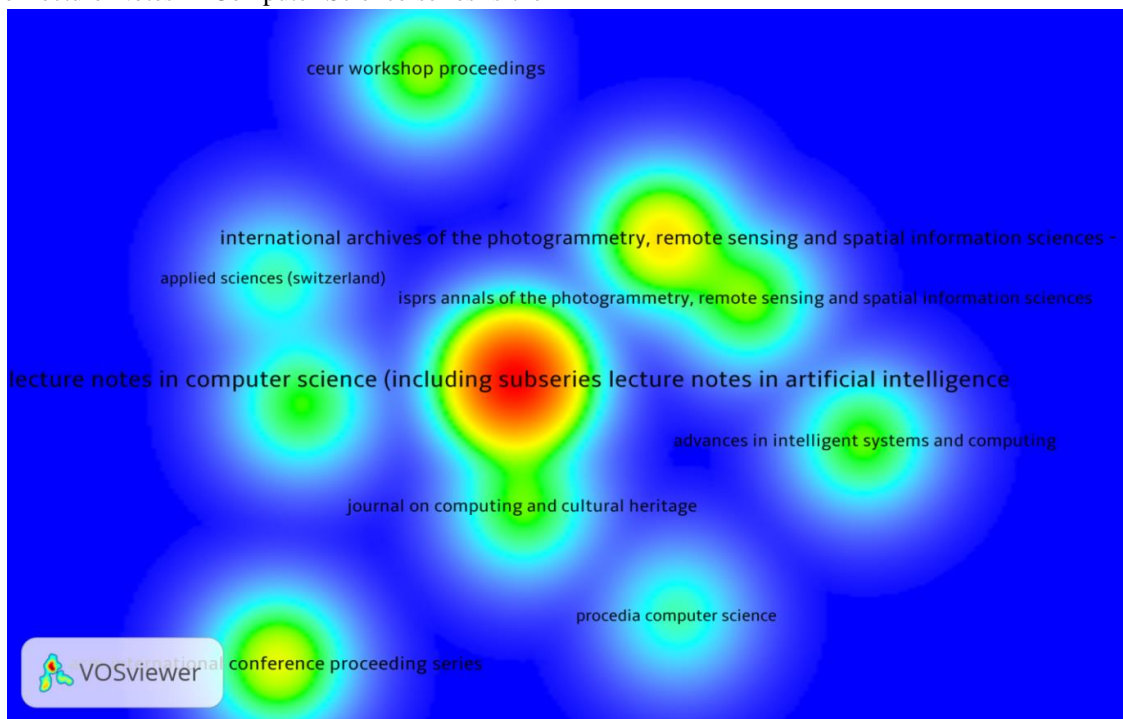


Fig. 4. Journals' bibliographic coupling (density visualization).

TABLE III. JOURNALS' BIBLIOGRAPHIC COUPLING

Source	Documents	Citations	Total link strength
“Lecture notes in computer science”	94	711	647
“Journal on computing and cultural”	13	396	429
“Applied sciences (Switzerland)”	7	28	263
“ACM international conference proceedings”	20	118	172
“ISPRS annals of photogrammetry, remote sensing, and spatial information sciences”	14	61	164
“Communications in computer and information science”	12	21	153
“International Archives of the photogrammetry, remote sensing, and spatial information sciences”	26	202	137
“Advances in intelligent systems and computing”	13	20	98
“CEUR workshop proceedings”	15	27	73
“Procedia computer science”	7	31	56

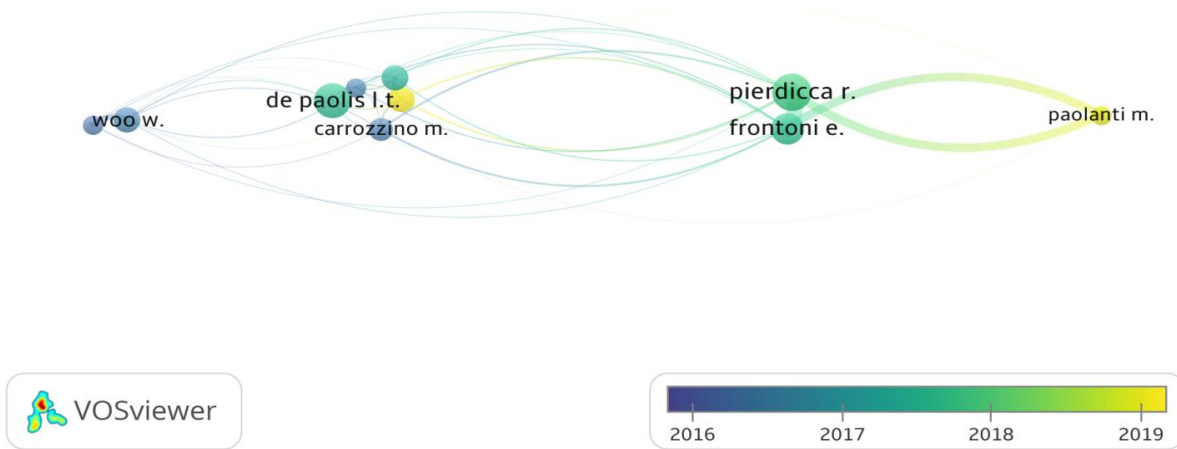


Fig. 5. Authors' bibliographic coupling (overlay visualization).

TABLE IV. AUTHORS' BIBLIOGRAPHIC COUPLING

Author	Documents	Citations	Total link strength
Pierdicca R.	12	339	804
Frontoni E.	10	337	798
Paolanti M.	6	46	423
Woo W.	8	66	161
Marto A.	8	34	143
Kim J.	6	54	138
Carrozzino M.	7	66	118
Liarokapis F.	8	290	94
Bostanci E.	6	66	87
De Paolis L.T.	11	214	62

With 12 publications, 339 citations, and a total link strength of 804, Pierdicca R. (Università Politecnica Delle Marche, Italy) was the most prominent author. With 10 articles and 798 links, Frontoni E. (Università Politecnica Delle Marche, Italy) is the second most cited author (337). Lucio Tommaso De Paolis (Università del Salento, Italy) is the second strongest author in terms of documents (11), citations (214), and link strength (62).

Liarokapis F. (Cyprus University of Technology, Cyprus) is the third strongest author in terms of citations (290), with 8 publications and 94 link strengths.

The first number reflects the total number of documents, the total number of citations are represented by the second number, and the total number of link strengths by the third number. The remaining authors are listed in order of total link strength: Paolanti Marina (University of Macerata, Italy) (6; 46; 423), Woo Woontack (Korea Advanced Institute of Science and Technology, Korea) (8; 66; 161), Marto Anabela (Polytechnic of Leiria, Leiria, Portugal) (8; 34; 143), Kim J. (Seoul National University, Seoul, Korea) (6; 54; 138), Carrozzino (6; 66; 87). There are five experts from Italy, two from South Korea, and one each from Turkey, Cyprus, and Portugal among the top ten authors.

It demonstrates Italy's dominance in this field of study. The color disparities in Fig. 5 depict the groupings of those writers

based on the years in which their research is published. It reveals that among these authors, Paolanti Marina, Pierdicca, and Frontoni E. have the most recent studies.

#### D. Bibliographic Coupling of the Publications

Fig. 6 shows the publications' bibliographic coupling with network visualization. The papers with at least 100 number of citations are considered for this study. Only 10 documents out of 590 met the criteria, and three of those do not have any coupling with other papers. As a result, just seven publications are studied. For each article, the number of citations and overall link strength is calculated.

The documents are arranged as per the total link strength as shown in Table V. Anderson et al. (2010) [57] has the highest link strength of 8 and the second highest citations of 205. Bekele et al. (2018) [58], on the other hand, have the most citations 240 and the second strongest link strength of 5. Styliani et al [2009; (207; 4)] [59], Wojciechowski et al [2004; (228; 3)] [60], Schmalstieg et al [2007; (115; 2)] [61], Tom Dieck et al. [2017; (101; 1)] [62], Sylaiou et al [2010; (146; 1)] [63] are the other papers listed in order of overall link strength.

The first number inside the parentheses represents the number of citations, while the second is the total link strength for each article.

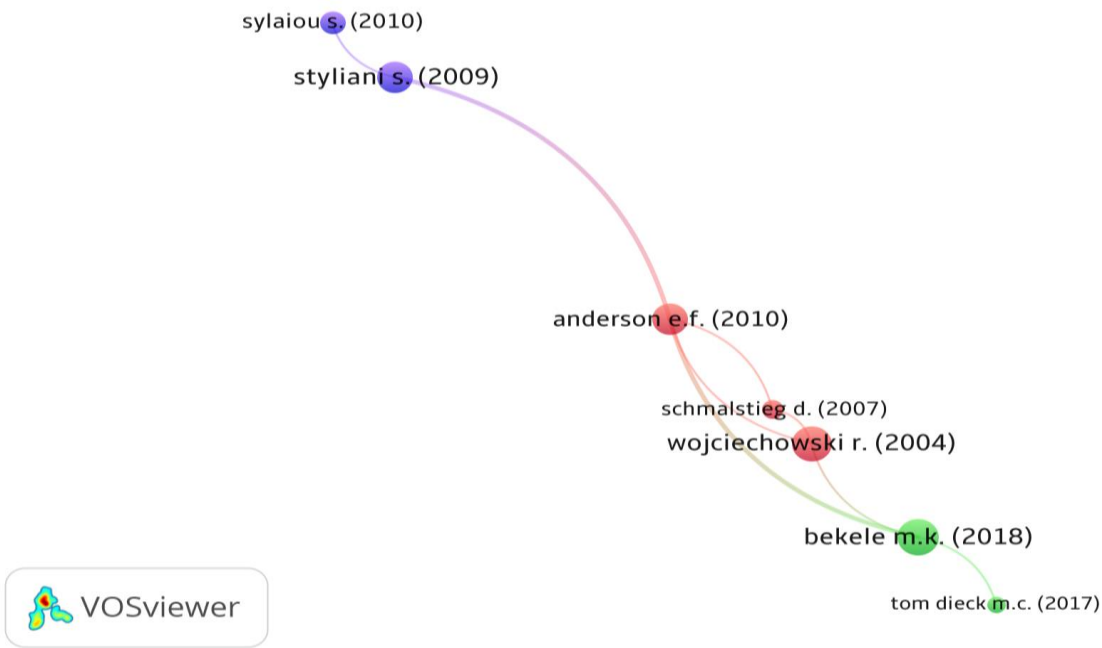


Fig. 6. Publications' bibliographic coupling (network visualization).

TABLE V. PUBLICATIONS' BIBLIOGRAPHIC COUPLING

Document	Citations	Total link strength	Domain of discussion
Anderson E.F.(2010)	205	8	Game Development for Cultural Heritage
Bekele M.K.(2018)	240	5	AR,VR and MR in Cultural Heritage
Styliani S.(2009)	207	4	Virtual Museums
Wojciechowski R.(2004)	228	3	AR, VR in museum exhibition
Schmalstieg D.(2007)	115	2	Handheld AR
Tom Diek M.C. (2017)	101	1	Visitor experience at heritage sites
Sylaious S. (2010)	146	1	Virtual Museum

E. Co-occurrences of the Author Keywords

Fig. 7 shows the author keywords' co-occurrences in a network format. The minimal number of occurrences of a keyword is 12 as an inclusion criterion. Only 14 of the 1296 keywords passed the test.

The frequency and its linkage with other keywords is computed for each keyword. Keywords with the most total link strength are chosen as shown in Table VI. With 332 occurrences and 377 total link strength, Augmented Reality is the most popular keyword. With 282 occurrences and 343 total link strength, Cultural Heritage is the second most popular keyword. Virtual Reality (91; 126), Mixed Reality (33; 67), Photogrammetry (23; 39), Storytelling (13; 30), 3D Reconstruction (14; 29), Education (16; 28), User Experience (15; 26), 3D Modeling (15; 23), Intangible Cultural Heritage (16; 20), Mobile Application (12; 20), Museums (12; 19),

Mobile augmented Reality (22; 17). For the rest, the first number inside the parentheses represents the number of occurrences, and the total link strength is represented by the second number.

Fig. 7 depicts different colored clusters representing the more frequently connected keywords. Education, Intangible Cultural Heritage, Mixed Reality, Mobile Augmented Reality, Narrative, and User Experience make up the largest cluster. The second cluster is comprised of augmented Reality, Cultural Heritage, Mobile Application, Photogrammetry, and 3D-Reconstruction. Each of the other three clusters consists of a single element each, viz. 3D-Modeling, Virtual Reality, And Museums. The research is mainly done in two broader areas. The first one is learning about cultural heritage through mobile applications using Mixed Reality. The second one is a 3D-reconstruction of cultural heritage using augmented Reality and photogrammetry.

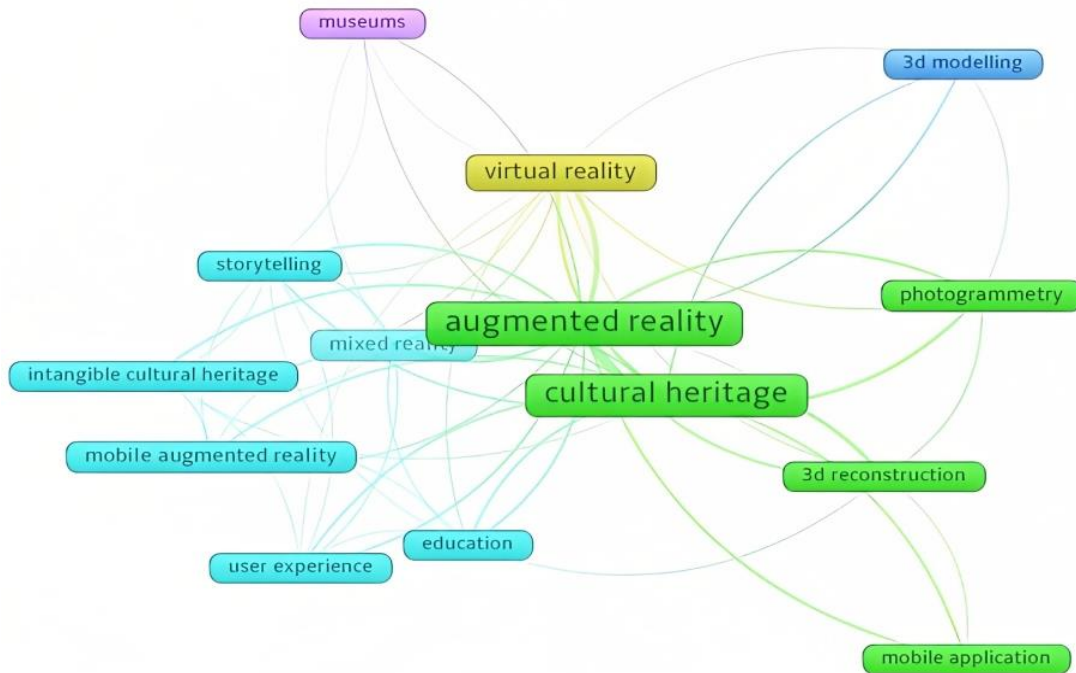


Fig. 7. The author keywords' co-occurrences (network visualization).

TABLE VI. THE AUTHOR KEYWORDS' CO-OCCURRENCES

Keyword	Occurrences	Total link strength
Augmented Reality	332	377
Cultural Heritage	282	343
Virtual Reality	91	126
Mixed Reality	33	67
Photogrammetry	23	39
Storytelling	13	30
3D-Reconstruction	14	29
Education	16	28
User Experience	15	26
3D-Modelling	15	23
Intangible Cultural Heritage	16	20
Mobile Application	12	20
Museums	12	19
Mobile Augmented Reality	22	17

## V. DISCUSSION

The following paragraph discusses how the research objectives specified in Section I.A, are achieved through bibliometric analysis and systematic literature survey.

1) Compile the peer-reviewed articles and conference papers on applications of AR/VR in cultural heritage during the period of 1999 to 2022.

In this work, publications from the Scopus database those are relevant to Augmented Reality or Virtual Reality-based solutions for learning and reconstruction of cultural heritage

are retrieved, analyzed, and visualized using descriptive and evaluative bibliometric analytic methodologies. To evaluate and visualize the linked data, VOSviewer 1.6.17 software is employed. Bibliographic coupling of publications, journals, countries, authors, and author keywords co-occurrences from associated publications are examined and illustrated.

2) Ascertain the spread of this field using bibliometric coupling of publications, journals, and authors.

As per the journals' bibliographic coupling, the premier journal in AR and VR literature related to cultural heritage is "Lecture notes in the computer science" series with the highest

publications, citations, and total link strength. Few more important journals in this area are; "Journal on computing and cultural heritage" (impact factor: 2.64), "International Archives of the photogrammetry, remote sensing and special information sciences" (impact factor: 0.82), "Applied sciences" (impact factor: 2.73), ACM International conference proceeding series (impact score: 0.61), "ISPRS Annals of the photogrammetry remote sensing and special information sciences" (impact score: 1.45), "Communications in Computer and information science" (impact factor: 0.48), "Advances in intelligent systems and computing" (impact factor: 0.63), CEUR workshop proceedings (impact score: 0.55), and "Procedia Computer science" (impact score: 2.09). The scopes of top journals are innovative usage of AR/VR for the discovery, reconstruction, experiencing, and learning of cultural heritage.

As observed from Table IV and Table V, the highest cited paper in this area is written by Anderson E. F. (2010) [57] which is co-authored by Liarokapis who is the eighth author as per the total link strength. Similarly, it is also observed that the second highest cited publication is by Bekele M. K. (2018) [58] and co-authored by Pierdicca R. and Frontoni E., who are the most cited authors. The third most cited publication is by Styliani S. (2009) [59] which is also co-authored by Liarokapis.

3) Understand which countries have the most publications in this area.

As per results shown in Table II related to the bibliographic coupling of countries, the maximum work in this area has been done in Italy which is the home of 58 UNESCO world heritage sites. The next three countries as per the number of publications are Greece with 18 UNESCO sites, the United Kingdom with 33 UNESCO sites, and Spain with 49 UNESCO sites. The other important countries have been: Portugal, France, Australia, South Korea, Germany, China, Malaysia, the United States, and Indonesia.

As per the authors' bibliographic coupling, the top three authors, Pierdicca, Frontoni E., and Paolanti are from Italy and worked together in many of the publications. Among the 10 strongest authors five are from Italy out of which the first two authors are from the same university that is Università Politecnica Delle Marche, Italy.

4) Which technologies were employed for implementation of cultural heritage projects?

As observed from Table I, most of the museum tour related projects are based on immersive technology and comparatively less work is done in non-immersive and semi-immersive areas. It is also found that mainly the immersive experience is provided by stimulating the optic senses of users, followed by the haptic, auditory, smell and taste senses.

The co-occurrences of author keywords show that the authors have largely employed augmented reality for 3D-reconstruction of cultural heritage using photogrammetry. It is also seen that Augmented Reality is used more often than Virtual Reality to provide users with an immersive experience of ancient cultural heritage.

By examining the co-occurrences of author keywords, the publications' and journals' bibliographic couplings, it can be

concluded that most of the researchers either tried to reconstruct the ancient environment of cultural heritage through mobile applications using Augmented Reality and photogrammetry or the researchers tried to create a submerging user experience of intangible cultural heritage through storytelling and using the Mixed Reality, especially for educational purpose. The most cited review papers by Anderson et.al (2010) [57] and Bekele et.al (2018)[58] also support this finding.

## VI. CONCLUSION AND FUTURE WORK

The results of the different types of analysis performed here indicate that the AR,VR and MR related research in the field of cultural heritage is mostly concentrated in Italy and surrounding European countries which are the home of many ancient monuments and other heritage sites. Other than Europe, some researchers from South Korea are also working in the same domain for providing information about their cultural heritage sites through mobile applications. It is also observed that although Mexico has 35 UNESCO heritage sites, the research in this area is still lagging behind there.

This study would help researchers to identify the prominent journals in the said domain, recognize stalwarts in the field and follow their works, find out path-breaking publications to refer to, predict the direction of future studies, etc. Here, we have presented the bibliometric analysis of the publications related to AR,VR and MR in cultural heritage. The work can be further extended to include publications related to specific areas such as 3D-reconstruction of heritage sites, gaming and educational applications related to this domain, etc. In a conclusion, it can be commented that AR, VR and MR in cultural heritage and 3D-reconstruction has a great scope and potential for immediate research.

## VII. CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

## REFERENCES

- [1] Mine TZ. Adaptive re-use of monuments "restoring religious buildings with different uses". Journal of cultural heritage. 2013 Jun 1;14(3):S14-9.
- [2] Loulanski T. Revising the concept for cultural heritage: the argument for a functional approach. International journal of cultural property. 2006 May;13(2):207-33.
- [3] Loulanski T. Cultural heritage and sustainable development: exploring a common ground. The Journal of International Media, Communication, and Tourism Studies. 2007;5:37-58.
- [4] Rambach J, Lilligreen G, Schäfer A, Bankanal R, Wiebel A, Stricker D. A survey on applications of augmented, mixed and virtual reality for nature and environment. InInternational Conference on Human-Computer Interaction 2021 Jul 24 (pp. 653-675). Springer, Cham.
- [5] Sirohi P, Agarwal A, Maheshwari P. A Survey on Augmented Virtual Reality: Applications and Future Directions. In2020 Seventh International Conference on Information Technology Trends (ITT) 2020 Nov 25 (pp. 99-106). IEEE.
- [6] Bergamasco M, Avizzano C, Di Pietro G, Barbagli F, Frisoli A. The museum of pure form: system architecture. InProceedings 10th IEEE International Workshop on Robot and Human Interactive Communication. ROMAN 2001 (Cat. No. 01TH8591) 2001 Sep 18 (pp. 112-117). IEEE.
- [7] Frisoli A, Jansson G, Bergamasco M, Loscos C. Evaluation of the pure-form haptic displays used for exploration of works of art at museums. InWorld haptics conference, Pisa, March 2005 Mar 18 (pp. 18-20).

- [8] Loscos C, Tecchia F, Frisoli A, Carrozzino M, Widenfeld HR, Swapp D, Bergamasco M. The Museum of Pure Form: touching real statues in an immersive virtual museum. In VAST 2004 Dec 7 (pp. 271-279).
- [9] Loscos C, Widenfeld HR, Roussou M, Meyer A, Tecchia F, Drettakis G, et al. The CREATE project: Mixed reality for design, education, and cultural heritage with a constructivist approach. In: Proceedings of the 2nd IEEE and ACM international symposium on mixed and augmented reality (ISMAR 2003). IEEE; 2003, p. 282-3.
- [10] Roussou M, Drettakis G. Can VR be useful and usable in real-world contexts? Observations from the application and evaluation of VR in realistic usage conditions. In: Proceedings of the 1st international conference on virtual reality. Mira Digital Publishing; 2005, p. 9.
- [11] Christou C, Angus C, Loscos C, Dettori A, Roussou M. A versatile large-scale multimodal VR system for cultural heritage visualization. In Proceedings of the ACM symposium on Virtual reality software and technology 2006 Nov 1 (pp. 133-140).
- [12] Inoue N. Application of ultra-realistic communication research to digital museum. In Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry 2010 Dec 12 (pp. 29-32).
- [13] Vi CT, Ablart D, Gatti E, Velasco C, Obrist M. Not just seeing, but also feeling art: Mid-air haptic experiences integrated in a multisensory art exhibition. *International Journal of Human-Computer Studies*. 2017 Dec 1;108:1-4.
- [14] Pursey T, Lomas D. Tate sensorium: an experiment in multisensory immersive design. *Senses Soc* 2018;13(3):354-66.
- [15] Obrist M, Boyle G, van Brakel M, Duerinck F. Multisensory experiences & spaces. In Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces 2017 Oct 17 (pp. 469-472).
- [16] Jozuka E. Multisensory cinema adds smell and touch to VR worlds. In: *Motherboard*. 2015, URL [https://motherboard.vice.com/en\\_us/article/qkv7pp/multisensory-cinema-adds-smell-and-touch-to-vr-worlds](https://motherboard.vice.com/en_us/article/qkv7pp/multisensory-cinema-adds-smell-and-touch-to-vr-worlds).
- [17] Butnariu S, Duguleană M, Brondi R, Gîrbacia F, Postelnicu CC, Carrozzino M. An interactive haptic system for experiencing traditional archery. *Acta Polytechnica Hungarica*. 2018 Jan 1;15(5):185-208.
- [18] Tennent P, Martindale S, Benford S, Darzentas D, Brundell P, Collishaw M. Thresholds: Embedding virtual reality in the museum. *Journal on Computing and Cultural Heritage (JOCCH)*. 2020 May 30;13(2):1-35.
- [19] Terracciano A. Zelige Door on Golborne Road: Exploring the Design of a Multisensory Interface for Arts, Migration and Critical Heritage Studies. In DHN 2018 (pp. 152-161).
- [20] Terracciano A, Dima M, Carulli M, Bordegoni M. Mapping Memory Routes: A multisensory interface for sensorial urbanism and critical heritage studies. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems 2017 May 6 (pp. 353-356).
- [21] Sardo JD, Pereira JA, Veiga RJ, Semião J, Cardoso PJ, Rodrigues JM. Multisensorial portable device for augmented reality experiences in museums. *International Journal of Education and Learning Systems*. 2018 Apr.
- [22] Rodrigues JMF, Ramos CMQ, Pereira JAR, Sardo JDP, Cardoso PJS. Mobile five senses augmented reality system: Technology acceptance study. *IEEE Access* 2019;7:163022-33.
- [23] Marto A, Melo M, Gonçalves A, Bessa M. Multisensory augmented reality in cultural heritage: impact of different stimuli on presence, enjoyment, knowledge and value of the experience. *IEEE Access*. 2020 Oct 20;8:193744-56.
- [24] Marto A, Melo M, Gonçalves A, Bessa M. Development and evaluation of an outdoor multisensory AR system for cultural heritage. *IEEE Access* 2021;9:16419-34.
- [25] McLaughlin ML, Sukhatme G, Shahabi C, Hespanha J, Ortega A, Medioni G. The haptic museum. In Proceedings of the EVA 2000 conference on electronic imaging and the visual arts 2000 Mar.
- [26] Garzotto F, Rizzo F. Interaction paradigms in technology-enhanced social spaces: a case study in museums. In Proceedings of the 2007 conference on Designing pleasurable products and interfaces 2007 Aug 22 (pp. 343-356).
- [27] Figueroa P, Boulanger P, Londoño E, Prieto F, Coral M, Borda J, et al. Multimodal exploration of small artifacts: An exhibition at the gold museum in bogota. In: Proceedings of the 16th ACM symposium on virtual reality software and technology (VRST '09). ACM; 2009, p. 67-74.
- [28] Osorio MF, Figueroa P, Prieto F, Boulanger P, Londoño E. A novel approach to documenting artifacts at the Gold Museum in Bogota. *Computers & Graphics*. 2011 Aug 1;35(4):894-903.
- [29] de Oliveira Lemos MO, Machado LS. New technologies to integrate haptics on Web: Expanding the Museu3I. In 2013 XV Symposium on Virtual and Augmented Reality 2013 May 28 (pp. 46-52). IEEE.
- [30] Chaudhuri S, Priyadarshini K. Cultural heritage objects: Bringing them alive through virtual touch. In Digital Hampi: Preserving Indian Cultural Heritage 2017 (pp. 337-354). Springer, Singapore.
- [31] Aniyath PK, Gopalan SK, Kumari P, Chaudhuri S. Combined haptic-visual and auditory rendering of cultural heritage objects. In Asian Conference on Computer Vision 2014 Nov 1 (pp. 491-506). Springer, Cham.
- [32] Popovici DM, Bogdan CM, Polceanu M, Querrec R. Applying of an ontology based modeling approach to cultural heritage systems. *Adv Electr Comput Eng* 2011;11(3):105-10.
- [33] Popovici DM, Polceanu M, Popescu A. Augmenting user experience in virtual environments through haptic feedback. In Proceedings of the 7th Balkan Conference on Informatics Conference 2015 Sep 2 (pp. 1-6).
- [34] Bie B, Zhang Y, Fu R. Study on display space design of off-line experience stores of traditional handicraft derivative product of ICH based on multisensory integration. In: Proceedings of the international conference of design, user experience, and usability. LNCS, vol. 10920, Springer; 2018, p. 459-70.
- [35] Ottoju K, Harrison S. Interaction as a component of meaning-making. In Proceedings of the 7th ACM conference on Designing interactive systems 2008 Feb 25 (pp. 193-202).
- [36] De Paolis LT, Aloisio G, Celentano MG, Oliva L, Vecchio P. Design and development of a virtual reality application for edutainment in cultural heritage. In 2009 15th International Conference on Virtual Systems and Multimedia 2009 Sep 9 (pp. 80-84). IEEE.
- [37] Paolis LTD, Aloisio G, Celentano MG, Oliva L, Vecchio P. A game-based 3D simulation of otranto in the middle ages. In: 2010 third international conference on advances in computer-human interactions. IEEE; 2010, p. 130-3.
- [38] Paolis LTD, Aloisio G, Celentano MG, Oliva L, Vecchio P. Experiencing a town of the middle ages: An application for the edutainment in cultural heritage. In: 2011 IEEE 3rd International conference on communication software and networks. IEEE; 2011, p. 169-74.
- [39] Paolis LTD, Aloisio G, Celentano MG, Oliva L, Vecchio P. A simulation of life in a medieval town for edutainment and touristic promotion. In: 2011 international conference on innovations in information technology. IEEE; 2011, p. 361-6.
- [40] Paolis LTD, Aloisio G, Celentano MG, Oliva L, Vecchio P. MediaEvo Project: A serious game for the edutainment. In: 2011 3rd international conference on computer research and development, vol. 4. IEEE; 2011, p. 524-9.
- [41] Tsuchida M, Yano K, Tanaka HT. Development of a high-definition and multispectral image capturing system for digital archiving of early modern tapestries of kyoto gion festival. In: Proceedings of the 20th international conference on pattern recognition. Istanbul, Turkey: IEEE; 2010, p. 2828-31.
- [42] Tanaka HT, Hachimura K, Yano K, Tanaka S, Furukawa K, Nishiura T, Tsutida M, Choi W, Wakita W. Multimodal digital archiving and reproduction of the world cultural heritage "Gion Festival in Kyoto". In Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry 2010 Dec 12 (pp. 21-28).
- [43] Faustino DB, Gabriele S, Ibrahim R, Theus A-L, Girouard A. SensArt Demo: A multisensory prototype for engaging with visual art. In: Proceedings of the 2017 ACM international conference on interactive surfaces and spaces. Brighton, United Kingdom: ACM; 2017, p. 462-5.
- [44] Hakulinen J, Keskinen T, Mäkelä V, Saarinen S, Turunen M. Omnidirectional video in museums – authentic, immersive and

- entertaining. In: Proceedings of the international conference on advances in computer entertainment. Springer, Cham; 2018, p. 567–87.
- [45] Guidi, G., Frischer, B., & Lucenti, I. (2007, July). Rome reborn-virtualizing the ancient imperial Rome. In Workshop on 3D virtual reconstruction and visualization of complex architectures. Trento, Italy: Fondazione Bruno Kessler.
- [46] Wood, Z.M., William, A., Yoon, A., Copeland, A. (2018). Virtual Bethel: Preservation of Indianapolis's Oldest Black Church. In: Ievenueberg, I., Neilson, T., Rheams, D. (eds) Research Methods for the Digital Humanities. Palgrave Macmillan, Cham. [https://doi.org/10.1007/978-3-319-96713-4\\_11](https://doi.org/10.1007/978-3-319-96713-4_11).
- [47] Foni, A., George Papagiannakis, and Nadia Magnenat-Thalmann. "Virtual Hagia Sophia: Restitution, visualization and virtual life simulation." In Proc. UNESCO World Heritage Congress, vol. 2. 2002.
- [48] Pritchard A. Statistical bibliography or bibliometrics. Journal of documentation. 1969;25:348.
- [49] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group\*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine. 2009 Aug 18;151(4):264-9.
- [50] Van Eck N, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. scientometrics. 2010 Aug 1;84(2):523-38.
- [51] Van Eck NJ, Waltman L. CitNetExplorer: A new software tool for analyzing and visualizing citation networks. Journal of informetrics. 2014 Oct 1;8(4):802-23.
- [52] Van Eck NJ, Waltman L. Systematic Retrieval of Scientific Literature based on Citation Relations: Introducing the CitNetExplorer Tool. InBIR@ ECIR 2014 Apr 13 (pp. 13-20).
- [53] Eck NJ, Waltman L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. Scientometrics. 2017 May 1;111(2):1053-70.
- [54] Kessler MM. Concerning some problems of intrascience communication. Massachusetts Institute of Technology, Lincoln Laboratory; 1958.
- [55] Kessler MM. An experimental communication center for scientific and technical information. Massachusetts Institute of Technology, Lincoln Laboratory; 1960.
- [56] Kessler MM. An experimental study of bibliographic coupling between technical papers. Massachusetts Inst of Tech Lexington Lincoln Lab; 1962 Mar 1.
- [57] Anderson EF, McLoughlin L, Liarokapis F, Peters C, Petridis P, De Freitas S. Developing serious games for cultural heritage: a state-of-the-art review. Virtual reality. 2010 Dec;14(4):255-75.
- [58] Bekele MK, Pierdicca R, Frontoni E, Malinverni ES, Gain J. A survey of augmented, virtual, and mixed reality for cultural heritage. Journal on Computing and Cultural Heritage (JOCCH). 2018 Mar 22;11(2):1-36.
- [59] Styliani S, Fotis L, Kostas K, Petros P. Virtual museums, a survey and some issues for consideration. Journal of cultural Heritage. 2009 Oct 1;10(4):520-8.
- [60] Wojciechowski R, Walczak K, White M, Cellary W. Building virtual and augmented reality museum exhibitions. In Proceedings of the ninth international conference on 3D Web technology 2004 Apr 5 (pp. 135-144).
- [61] Schmalstieg D, Wagner D. Experiences with handheld augmented reality. In 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality 2007 Nov 13 (pp. 3-18). IEEE.
- [62] Jung TH, tom Dieck MC. Augmented reality, virtual reality and 3D printing for the co-creation of value for the visitor experience at cultural heritage places. Journal of Place Management and Development. 2017 Jun 5.
- [63] Sylaiou S, Mania K, Karoulis A, White M. Exploring the relationship between presence and enjoyment in a virtual museum. International journal of human-computer studies. 2010 May 1;68(5):243-53.

#### AUTHORS' PROFILE



Nilam Upasani is working as an Associate Professor in the Dept. Of Computer Engineering at Pimpri Chinchwad college of Engineering, Pune, India. She received her Bachelor and Master degree in Computer Engineering from Pune University, India. She received her Ph.D. in Computer Science and Engineering from IIT (ISM), Dhanbad, India. She has published several research papers in International Journals and presented many at International conferences of high repute. Her research interests include Data Science, Machine learning, Biomedical applications, and rare-class detection.



Asmita Manna is working as an Assistant Professor in the Dept of Computer Engineering, Pimpri Chinchwad College of Engineering. She obtained her Ph. D degree from Jadavpur University, Kolkata, India. Prior to that She obtained her Master Degree in Software Engineering from Jadavpur University, Kolkata, India and Bachelors Degree in Information Technology from Bengal Engineering and Science University, Shibpur (presently known as IEST, Shibpur). She has published multiple research papers in International Journals and presented at many conferences of international standards. Her research interests are Cyber Security, Information Systems, Data Science and Machine Learning.



Manjiri A. Ranjanikar received the B.E. and ME degree in Computer Science and Engineering from MGM's COE, Nanded and PhD. degree from Department of Computer Science and Engineering, SGGGS Institute of Engineering and Technology, Nanded, Maharashtra state, India in 2010, 2013, and 2020 respectively. Currently she is working as Assistant Professor in the Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, MH, India. She has published about 15 papers in international and national journals and conferences and 3 patents. She has reviewed IEEE Conference papers and book chapters. She is having 11 years of experience in teaching. Her areas of interest are Image processing, machine learning, Computer Vision, and Pattern classification.

# Implementation of Business Intelligence Solution for United Airlines

## Business Insights and Data Analytics for United Airlines Industry

Ng Iris, Sarasvathi Nagalingham

Faculty of Data Science and Information Technology, INTI International University, Nilai, Negeri Sembilan, Malaysia

**Abstract**—US Airline is recognized as the world's largest airline, with a massive number of daily departures completed and a combined fleet of over 2700 aircraft. US Airlines have carried major 18 airlines, categorized as mainline, regional, and freight airlines. United Airline is one of the major airlines in the US after American Airlines and Delta Airlines in the world. Today, companies received as much feedback from their customers. Customers can share their opinion and emotion through social media platforms, such as Twitter. Thus, collecting and understanding customer's opinion become the key benefits for the aviation industry to get actionable insights while increasing their competitiveness. Such insights are useful in planning and execution to increase the relationship with customers. Thus, this study was conducted to analyze customer's feedback in different airlines to discover actionable insights that increase the competitiveness of United Airline. The analysis result will be visualized on Tableau dashboards and BI solutions will be provided. By implementing the BI solutions, United Airline can make accurate decisions and define next strategies by identifying those positive and negative references. Thus, United Airline can improve the quality of their service, enhance customers loyalty, and boost business profitability.

**Keywords**—Business intelligence; aviation industry; dashboard visualization; tableau; data analytics

### I. INTRODUCTION

The COVID-19 pandemic is having a significant impact on the aviation industry, including a sharp decline in demand for passenger air travel. Demand for airline services has fallen dramatically because of changes in passenger behavior brought on by the COVID-19 crisis and travel restrictions. Additionally, containment efforts are endangering the viability of numerous businesses in the aviation industry as a whole, putting many jobs at risk. Despite some setbacks from the COVID-19 wave, the herd vaccination campaign is now beginning to pay off. Many nations have carried out their plans to resume some semblance of normality by removing numerous health restrictions, easing travel stipulations, and reopening borders. The momentum generated by the reopening had a positive and immediate impact on global demand for air travel.

Big data insights can provide airlines with a significant competitive advantage. The airline industry can learn a lot about their customers during the booking, check-in, boarding, and even during the flight. Airlines, along with loyalty programmes, generate arguably more customer data than any

other industry. This data contains vast amounts of valuable intelligence that affects operations, efficiency, and service.

### A. Business Intelligence Methodology

Business intelligence is a collection of technologies that are used to help businesses make better use of data to enhance decision-making quality. It is a process of extracting a large amount of data from information and knowledge to provide accurate intelligence on the nature of the market [1]. With business intelligence, the term business analysis is often used interchangeably. Business analysis, however, is a subset of business intelligence because business intelligence addresses strategies and instruments, while business analysis focuses more on techniques. Business intelligence is descriptive, whereas forecasts for business analysis are indicative and used to solve problems or business issues [2].

Business intelligence tools are application software that can be used to gather and process vast amounts of unstructured data. With the BI tools, it can also help the business to prepare data information for analysis, making it easy to build organizational reports and create dashboards to make visualizations of the information. As a result, decision-making can be accelerated and improved by staff and managers, operational efficiency improved, new revenue potential clarified, market trends identified, true KPIs reported, and new business opportunities identified [3]. There are many types of business intelligence tools such as Tableau, Power BI, Yellowfin BI, and more [4]. They can perform data analysis by generating reports, summaries, dashboards, maps, graphs, and charts that are informative and easy to understand.

### B. Overview of United Airline

US Airline is also recognized as the world's largest airline, with a massive number of daily departures completed and a combined fleet of over 2700 aircraft. There are flights departing and arriving every day from different countries. There are almost 889 million passengers being transported in 2018, according to the research, which can be recognized as the largest air travel market in any single country. US Airlines have carried major 18 airlines, categorized as mainline, regional, and freight airlines. United Airline is one of the major airlines in the US whose headquarters are located at Willis Tower in Chicago. United Airlines is as known as the third largest airline after American Airlines and Delta Airlines in the world. The slogan of United Airlines is "Fly the Friendly Skies" [5].



The main business goal of United Airline was to provide the highest quality of service for every single passenger wherever they encounter the airline. Moreover, to create repeat customers through satisfying them on their leisure or business travel. To achieve this, United Airlines had to provide top quality service which must deal with the customer in a respectful and helpful way. Furthermore, ensure their service meets customer needs by exploring the new alternatives and goods for enhancing customer loyalty [6].

Since the United States is a prosperous country, there are also many people traveling between countries. Therefore, many customers book or buy air tickets online or at the counter every day. United Airlines must collect the associated information to understand the daily flights, customer purchase records, and so on. Through the data collected, United Airlines can improve the quality of service, improve customer loyalty, meet everchanging customer needs, and boost business profitability.

### C. Challenges Faced by United Airline

- **Stiff competition:** The aviation industry has the characteristics of high investment, high risk, high technology, and low profit. In the United States, there are a lot of airline companies such as United Airlines, Delta Airlines, Southwest Airlines, Virgin Airlines etc. In the airline industry, demand for air travel and capacity has seen tremendous growth, but profitability is comparatively poor, even though unit costs have decreased by half in the past 40 years. Much of the productivity benefits from reduced prices have been passed on in the form of lower airfares to consumers [7].
- **Social media:** Over the past few years, social media has changed the Internet completely. Social media creates a bigger and easier forum for consumers to communicate and share knowledge and opinions. The content discussed and posted on social media can come in many shapes and forms, such as text, pictures, audio, and video, which can be used to communicate different types of messages for different scenarios. Airlines also receive airport and flight experience complaints from travelers. In addition to formal grievances, individuals nowadays often express views on social media about their travel experience, especially when unsatisfied. Such feedback regarding the preferences of customers is highly revealing. The knowledge is used by advanced data analytics to identify patterns and best practices in customer service.
- **Digital transformation:** Customer data, product data and operational data are required for the aviation company to stand out from competitors. United Airlines can collect thousands of operational and transaction data every day from various departments. Thus, how to analyze the data is important to get the information insights and discover hidden knowledge. The data itself has no value what matters what United Airlines is do with the data. Poor data quality such as missing values, noisy data and outliers may become a big challenge for the United

Airlines when processing with [8]. Without general knowledge of the various methods of data processing, incorrect information can be presented. It can be very risky for decision makers without the aid of a trained data scientist to analyze the data. Therefore, several prejudices can exist, and their decision making cannot be validated by the evidence.

### D. Opportunities of United Airline

- **Enhance customer experience:** Companies that truly recognize and prioritize their consumer demand benefit from crucial competitive advantages such as higher revenue and market shares and more loyal customers. They can distinguish themselves and improve customer loyalty from their rivals. Introducing new channels to their customer service, such as chatbots on websites, or event-trigger customized deals or communications, the customer experience is enhanced by data analytics. These tools detect trends based on previous experiences and provide clients with optimized solutions to their most common problems, leaving human interaction to solve more complicated problems [9].
- **Improve customer loyalty:** Airline companies get access to detailed information about each individual customer. They manage to obtain the customers' recent social media interactions with them and even access to customers' previous journey. With the help of BI, airline companies can get a holistic view of each customer. More specific marketing communications can be offered by airlines companies and personalize customer experience. This eventually improves the conversion of campaigns and consumer satisfaction. Hence, retain the customer loyalty [9].
- **Reduce marketing expenses:** Among the various airline companies, it can be competitive for the customer to choose the right airline for them. It might depend on different factors, such as overall ratings for the airline company, the seat cleanliness of the airplane, the food and beverage that was provided on plane and the value for money. A customer would choose their preferred airline company over another. Some might choose United Airlines because of the tidiness of the seats. With the help of BI solution, the company would be able to analyze the preferences of each customer and get to know how these factors affect the customers and give impact to the company. The marketing expenses could be reduced, and more revenue could be used on other sectors.

## II. DATASET ANALYSIS

The dataset is grabbed from Kaggle which is the Twitter airline sentiment. The original source version is slightly reformatted from Crowdfower's Data for Everyone library [10]. Both a CSV and SQLite files are included with it. The dataset basically contains the sentiment of tweets for 4 airline companies, which are Delta Airline, Southwest Airline, United Airline and Virgin Airline. The Table I and Table II identify the description of each attribute for two different datasets.

TABLE I. DATASET ATTRIBUTES AND DESCRIPTION: TWEET

No.	Attribute Name	Description
1	tweet_id	The unique ID of this tweet
2	airline_sentiment	The customer's opinion whether it was positive, neutral, or negative
3	negativereason	The reason why customers give negative opinions or feedback
4	airline	The name of the airline
5	name	The username of the Twitter account
6	retweet_count	The number of Twitter retweets
7	text	User's feedback
8	tweet_coord	The coordinate that the tweet has been created
9	tweet_created	The date the user posted the tweet
10	tweet_location	The tweet location of the user
11	user_timezone	The time zone of the user

TABLE II. DATASET ATTRIBUTES AND DESCRIPTION: AIRLINE REVIEWS

No.	Attribute Name	Description
1	airline	The name of the airline
2	overall	The overall rating
3	author	Passenger name
4	review_date	Date that the customer review
5	customer_review	Passenger review on their flight experience
6	aircraft	The type of aircraft
7	traveller_type	The purpose of the customer's flight
8	cabin	The cabin class of accommodation on a passenger ship
9	route	Flight path of the airline
10	seat_comfort	Passenger ratings for inflight seat comfortability
11	cabin_service	Passenger ratings for inflight cabin service
12	food_beverage	Passenger ratings for inflight food and beverage
13	entertainment	Passenger ratings for inflight entertainment
14	ground_service	Passenger ratings for ground service
15	value_for_money	Whether the travel is value for money or not
16	recommended	Whether the passenger will recommend to their friends or not

### A. Dataset Cleaning

Microsoft Excel is a software that can be used to store, manipulate, and present the data sets. It provides various types of tools to make sense of the datasets using its characteristics and formulas. Thus, Microsoft Excel will be used to clean the datasets, so that noisy data and duplicate values can be identified and removed.

- **Remove duplication:** Large datasets appear to have redundant material. Identifying duplicate data and deleting it is the first step in data pre-processing. Duplicate data has given us little benefit. The performance and quality of the dashboard will be reduced. Thus, the duplicate records are removed from the datasets. There are a total of 70711 duplicate records have been removed. Now the datasets remain 61185 unique values.

- **Split texts into columns:** To perform text analysis using word cloud, it is necessary to split the whole text into each individual column. Thus, select the entire column of the text then split text to column with delimiters. It is required to select the delimiter type to split the text into columns. The selected delimiters include semicolon, comma, space, and full stop.

### III. BUSINESS INTELLIGENCE ARCHITECTURE

Business intelligence architecture refers to the architecture used by an organization to define their data source of collection information processing and all the technology that supports their business intelligence. BI architecture consists of various parts and layers and every single of them are important as each of the components have its own purpose [11]. The components include:

- **Data collection:** The very first steps of business intelligence. Nowadays data is becoming more and more important to drive a business to success. Depending on a company's criteria and finances, a company may obtain their data through different data sources, such as CRM, ERP, databases, files, or APIs. These applications and methodologies collect data from internal processes and external sources of organisations.
- **Data integration:** The following step will be data integration which uses the data collected previously and proceeds on with the Extract, Transform and Load (ETL) process. For extract, the data will be pulled out from external sources. Then transform the data into a demand standard which also ensures the data is clean with no noise data. After that, load it into the destination data warehouse.
- **Data warehouse:** This is the place where all the data is being stored. The data warehouse outlines data for upstream applications in dimensions and fact tables. Processes performed in a data centre are data cleansing, metadata management, data distribution, resource management, recovery, and backup planning.
- **Data analysis:** This phase will be focused on data processing after it has been handled, transformed, and cleaned in previous steps by the data warehouse. The ubiquitous need for effective analysis to enable businesses of all sizes to grow and benefit is accomplished through BI application tools. Modern BI tools like Tableau enable users to create queries using drag and drop, building visualization graphs within a few clicks.
- **Data distribution:** When it comes to sharing information and providing stakeholders with indispensable perspectives to achieve sustainable business growth, data distribution is one of the most critical processes. distributed can be performed in 3 ways which are reporting, dashboard and embedding. Reports can be viewed and shared among recipients on a defined schedule. Alternatively, a dashboard can be shared for the recipient to view but not able to make changes. Without the need to mark the BI tool on external

applications of intranets, your own application may use dashboards as a means of analytics and reporting.

- Actionable insights: Data driven decision will be the final stage for the BI architecture. With the insight taken from the previous stage, higher management like CEO will have the ability to produce true, specific, data-based decisions that will assist them to move forward. This helps a high-level manager to get an understanding of strategic growth and future decisions to build a profitable organisation. Fig. 1 shows the example of BI architecture.

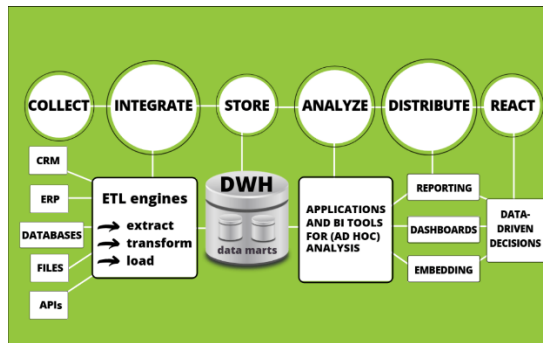


Fig. 1. Example of BI architecture.

#### IV. DIMENSIONAL MODEL

A dimensional model is a data structure consisting of a table of facts and tables of measurements. A fact table consists of measures that allow us to quantify quantitative data with ease of dimensional modelling. The dimension tables include descriptions of the fact table that surrounds the business process case. The following section discusses five steps to construct a dimensional model [12].

- Identify business process: Business process must be identified because it needs to define how the data stores in the data warehouse. The way to store the data should be efficient to process the data analysis. The data is scraped on the Internet and compared between the different companies. It could be used to identify the pros and cons of the different airline companies and help the company in terms of their decision making on marketing purposes.
- Identify granularity: This step is to identify the granularity of each table of facts and the business procedure. It defines the fact and dimensions table during this process. The dataset should include daily monthly quarterly granularity in a table.
- Identify dimensions: The time dimension must be specified because to know when the metrics which have been reported. The dimension would be the user, airline, tweet, location, and date. The date would have the granularities such as month, week, and day.
- Identify facts: A fact table contains information about measurements and foreign keys to dimension tables. The fact table includes numerical measurements, retweet\_count, airline\_sentiment\_confidence, negativereason\_confidence, overall, seat comfort, foodbev, valueformoney,

foodbev, value for money are the measurements for this fact table. The foreign keys would be the primary keys of each dimension.

- Build schema: Star schema is chosen for this research because it is easy to implement. The dimension tables are joined with the fact table with a common key attribute. Each dimension is highly denormalized in a single table. This could allow the system to optimize the query processing time. The schema can be implemented on the data warehouse with the help of ETL process which processes dimensional modelling. Fig. 2 shows the star schema developed for United Airline.

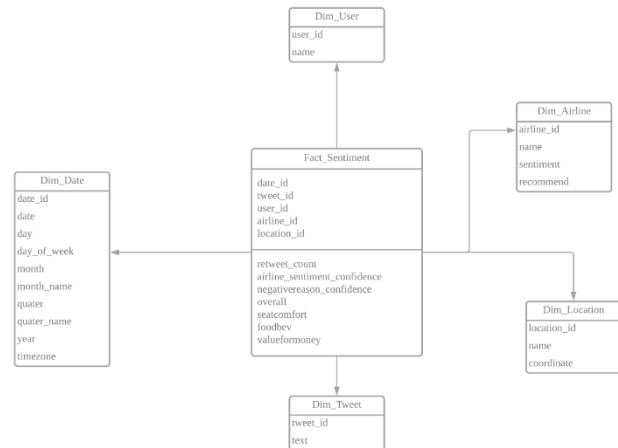


Fig. 2. Star schema for united airline.

#### V. DASHBOARD VISUALIZATION

There are 4 dashboards created for this study, Airline Recommendation Analysis, Passenger Review Analysis, Customer Sentiment Analysis, and Airline Rating Analysis. Fig. 3 section describes the dashboards in detail.

##### A. Airline Recommendation Analysis

The airline recommendation analysis is a dashboard which provides insights from different perspectives. There are four major airline companies which are Delta Airlines, Southwest Airlines, United Airlines, and Virgin America. This dashboard is focused on the recommendation of the users whether the users would recommend the particular airline which they have tried before, this could be critical because it might affect the customer's loyalty and company's profits. It allows the users to have a better understanding of the overall recommendation of each airline company. Moreover, it provides some drill-down features about the airline companies and the cabin types.

Fig. 4 shows the total percentage of users who are recommended and not recommended to the airline companies. So, from the pie chart below it shows that the users who have taken flight before do not recommend their airline companies which they have taken before. This might be related to several reasons, such as food and beverage availability in the airplane, the comfortless of the seat on the plane, and the value for their money.

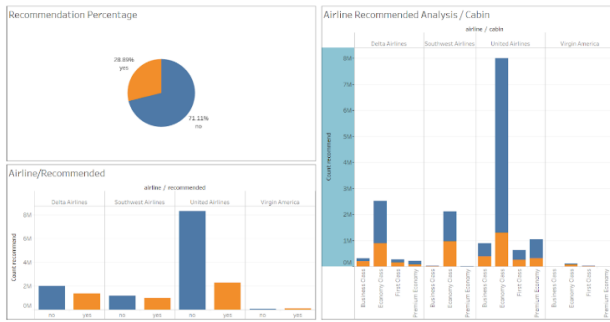


Fig. 3. Airline recommendation analysis.

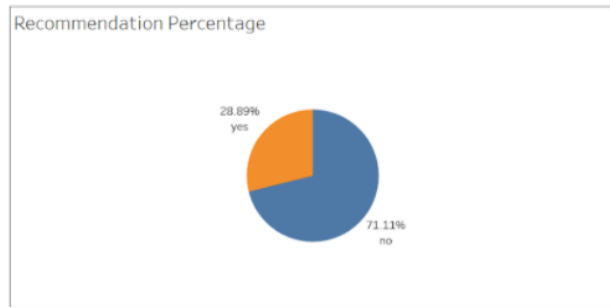


Fig. 4. Recommendation percentage.

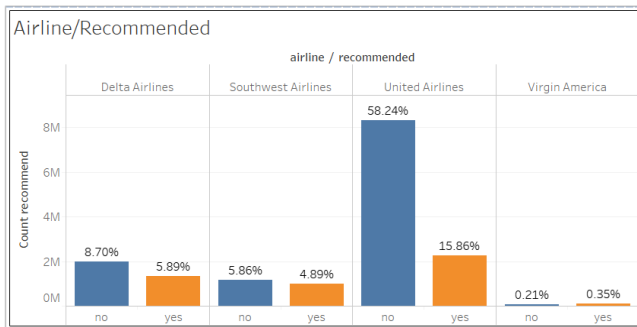


Fig. 5. Airline recommendation.

Fig. 5 shows the number of users who recommend and not recommend the airline company. Although Virgin America has the least number of records which has a total of 0.56%, most of the customers who had travelled with it were willing to recommend the airline to others. The company United Airlines has the highest number of customers who had travelled with them before, but most of the customers do not recommend it to others which consists of 58.24% among all the airline companies. From the customer perspective, it indicates that the United Airlines does not do a great job to serve their customers and it might be better to choose the other airlines rather than United Airlines. Although there are many customers who do not recommend United Airlines, there are still many customers who have taken flight with them. The reason behind this might be because of the cost of the air ticket since most people only care about the cost and want to keep the cost to the lowest. Fig. 6 stacked bar chart shows the drill-down features with the airlines and cabin types. The users would have a clear understanding of each cabin type and how they relate to the number of customers. In the economy class of the cabin type, all the four airline companies own the highest number of passengers. The reason for this might be

that the cost for the air ticket is the cheapest and most value of money when compared with the other cabin types. Among the economy class passengers, it indicates that most customers are from United Airlines which is true from the previous section, because it sold the most air tickets among his competitors.

**B. Passenger Review Analysis**

Fig. 7 shows dashboard for passenger’s review analysis that is able to provide useful insight to the airline company. The overall review of cabin class can give insight of which type of cabin class should be improved more to meet passenger’s satisfaction. The number of reviews shows the total number of passengers reviewed from the year of 2014 to 2018. Count of cabin will define the cabin type taken for the passengers. Lastly the word cloud of review will display the most frequent word collected from passengers’ review. The dashboard will be split and explained in more details in following page.

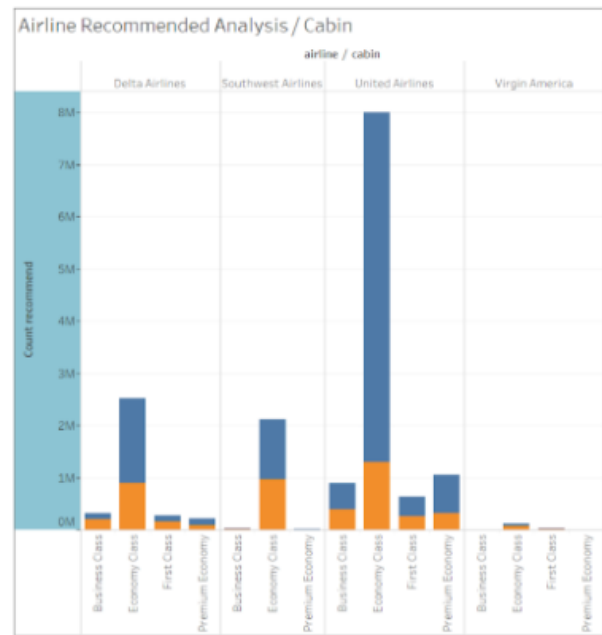


Fig. 6. Recommendation cabin.

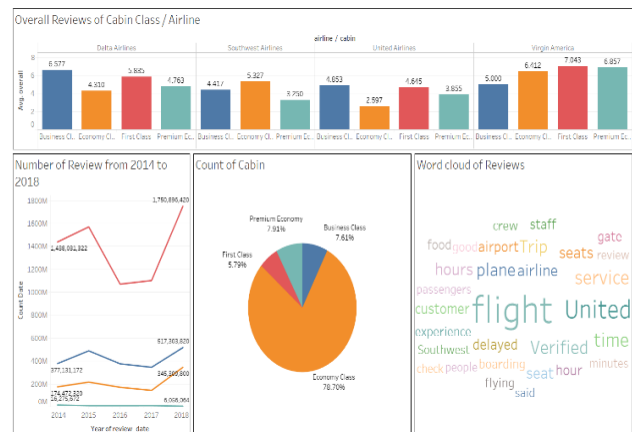


Fig. 7. Passenger review analysis.

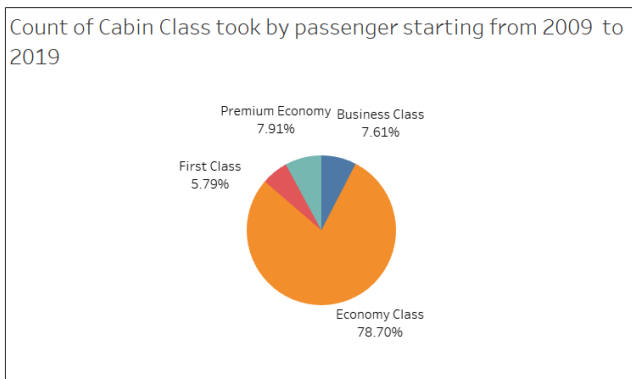


Fig. 8. Count of cabin class.

In addition to that, the Fig. 8 shows the total number count of airline cabin classes of all flights from the year of 2009 to 2019. There are 4 types of cabins provided for the passengers which are Economy class, Premium Economy, Business Class, and First-Class cabin. According to the diagram, it is obvious that most of the passengers took the economy class cabin which consisted of 78.70% and overwhelming other types of class. Business Class and Premium Economy class are very close to each other where only have a difference of 0.3%. However, Premium Economy cabins manage to get the second highest count with a percentage of 7.91%. Lastly, First Class cabins only have 5.79% of count due to the high price and luxury service which is not affordable by normal passengers.



Fig. 10. Word cloud of reviews.

Other than that, the Fig. 10 shows the reviews from the customer are important for the airline company as they can obtain information from it and make improvements. However, there are more than millions of reviews and the time taken for the airline company to read one by one will be a waste of time and effortless. This is the place where word cloud plays an important part as it will display the most relevant word from the reviews. Figure above shows the top 30-word from the customer. The larger the word the higher the count of the word. The largest word that observed from the diagram above was “flight” which means the reviews are related to flight. Furthermore, the author noticed that “delayed”, “hours” and “minutes” appear on the word cloud which mean that there are many reviews related to the flight delay which can up from minutes to hours. Another word “service”, “good” and “experience” can be concluded as the passengers’ experience with the good service.

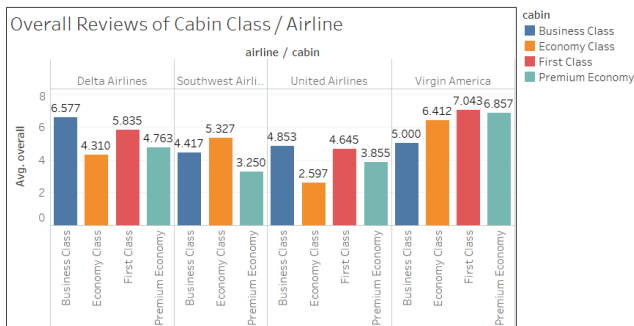


Fig. 9. Overall reviews of cabin class.

Apart from that, the Fig. 9 shows the overall reviews of the cabin types in each airline. There are four types of cabins which are business class, economy class, first class and premium economy. In the United Airlines, the economy class is the lowest selling ticket which only consists of 2.597 in average. This could lead to some serious problems because most of the seats in the airplane are economy class and if the more seats are empty, United Airlines would not be able to obtain the highest profitability on each flight. To get the highest profitability in terms of efficiency and effectiveness, the ticket sold on each cabin class should be equally distributed which would result in the plane being filled with customers in each flight.

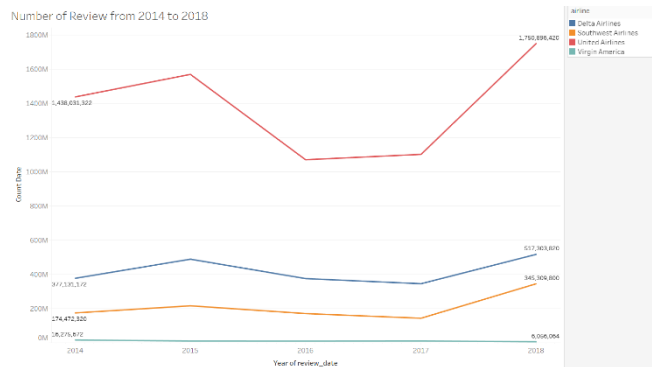


Fig. 11. Number of reviews.

Furthermore, the Fig. 11 shows the number of reviews that the airlines companies received from 2014 to 2019. There are four airline companies related in this diagram which are United airline, Delta airline, Southwest airline, and Virgin America airline. The number of reviews received for Virgin America was the lowest when comparing to other airline companies. In the year 2014 it only had 16,275,672 times of review and it encountered small drops steadily all the way to the year of 2018 until 6,056,064 times of review. For Southwest and Delta Airline they encounter slight upward and downward from 2014 to 2018. Both airlines were increased from 174,472,320 to 345,309,800 and 377,131,172 to 517,303,820, respectively. For United airline, it had the greatest number of reviews, and it is overwhelming the other 3 airlines. In the beginning of 2014, it only had 1400 millions of reviews while in 2018, United airline reviews rose to a new high of 1750 million.

C. Customer Sentiment Analysis

In this section, the customer opinions as well as emotions towards the services that the airline companies provided will be discussed in detail. This dashboard provides information from a different viewpoint. Sentiment analysis, also known as opinion mining, helps airline companies to analyze the sentiment of a piece of text and present it in the form of a dashboard. Basically, the study of text sentiment helps us to get an understanding of whether a piece of text is positive, negative, or neutral. While at the time the airline companies gathered the tweets, this gives them a clear sense of the public's opinion about the service they provided and tells the airline companies anything about what exactly people were talking about positively or negatively. Sentiment analysis will tell the airline companies what a piece of text's sentiment is. But people-generated text typically talks about more than one thing and sometimes has more than one emotion. Someone might write, for instance, that they didn't like the customer services of the United Airlines, but they liked the inflight experiences and feel comfortable on the journey, and a model of document-level sentiment analysis would only look at the whole document and sum up whether the overall feeling was mostly positive or negative.

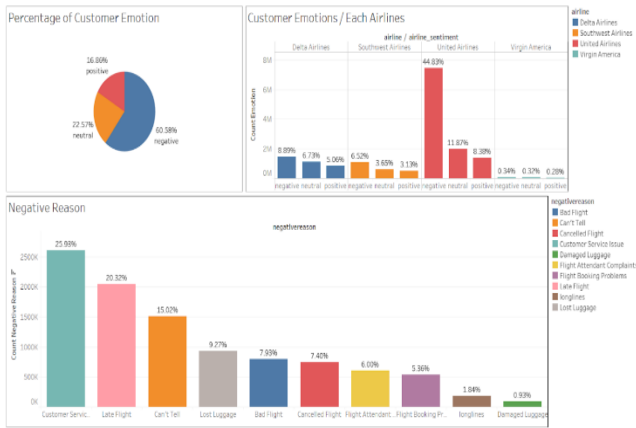


Fig. 12. Customer sentiment analysis dashboard.

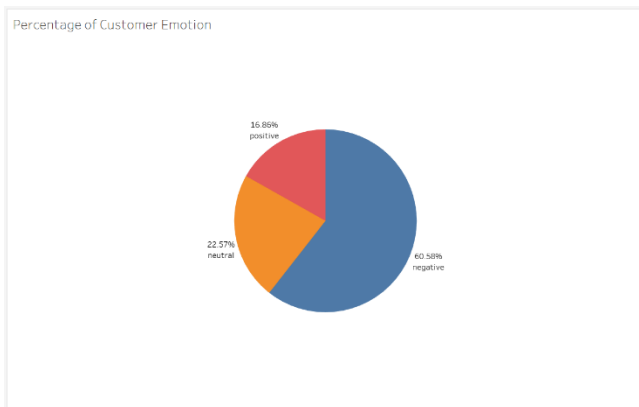


Fig. 13. Customer emotion.

Customer satisfaction plays an important role within the airline company (Fig. 12). It is not only the leading metric for assessing consumer satisfaction, recognizing dissatisfied customers, minimizing turnover, and increasing revenue, but it

is also a crucial point of differentiation that allows you in competitive market environments to gain new customers. Thus, airline companies need to analyze and understand the customer's experience to improve the services and products provided. Fig. 13 shows the percentage of customer's emotion about the flight. Based on the result, we can conclude that most of the passengers were not satisfied with the services provided by the airlines company. More than 60% of the passengers write a negative response to the airline companies, while only 16.86% of passengers feel satisfied with the services provided by the airlines company. While 23% of the passengers did not specify their emotions on the flight experience. Since most of the passengers were not satisfied with the service provided by the airline companies, thus the airline companies need to figure out the reason why the customer's feel bad about the flight. Else, the companies may lose a lot of customers and profit from that.

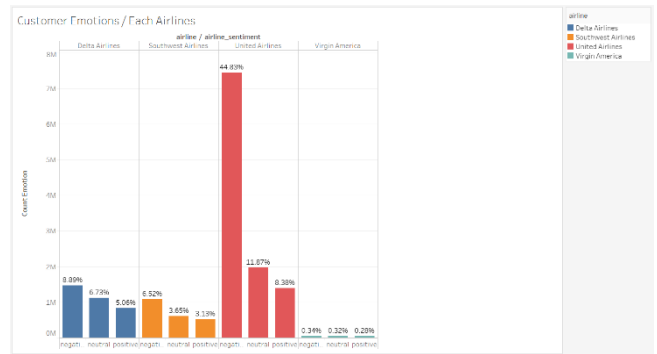


Fig. 14. Customer emotion on each airline.

Since the overall sentiment of customers towards the U.S. aviation industry has been analyzed, drill down is also necessary to know the customer's emotion on the specified airlines companies, so that we can compare the customer's reviews of United Airlines and its competitors, which are Delta Airline, Southwest Airline, and the Virgin Airline. The Fig. 14 shows the customer emotions in each airline. There are three types of emotions which are positive, negative, and neutral. In the United Airlines, 44.83% of customers have negative emotions about the services provided, while around 12% of passengers did not express their emotions, and only 8.38% of the passengers feel satisfied with the services provided by the United Airlines. However, if we look at the other airline companies, the gap in customer satisfaction with service is not very large. This result could lead to some serious problems which may result in loss in both customers and profit, since the customers are more likely to pursue better services. Thus, if the United Airlines did not figure out why customers were not satisfied with the services provided, it is possible for the customers to switch to the competitive airlines.

As the Fig. 14 analyzed how many people feel unsatisfied with the service provided by the airline companies, it is important for the airline companies to look at the negative reviews to find out the reason. The Fig. 15 shows the number of negative reasons for each airline. The negative reasons are basically classified in several types, such as customer service issues, flight cancelled, luggage lost and damaged, flight

booking problems, longlines, etc. Among them, around 26% of passengers did not like the customer services, while around 20% of the passengers feel unhappy because of the flight delays. Around 10% of passengers unhappy because their luggage lost and damaged during the transportation, while 7.40% of passengers unhappy because their flight was cancelled. These are the major reasons which tell the airline companies why the passengers did not satisfy the airline service. Remaining 20% of the passengers feel unhappy because the booking system is not efficient, or unhappy because of the longlines which required them to waste a lot of time booking for a flight and waiting for the flight. However, 15% of passengers did not want to tell why they feel unhappy on the service. Thus, it is important for the airline company to know in detail which service needs to be improved to fulfil the customer needs.

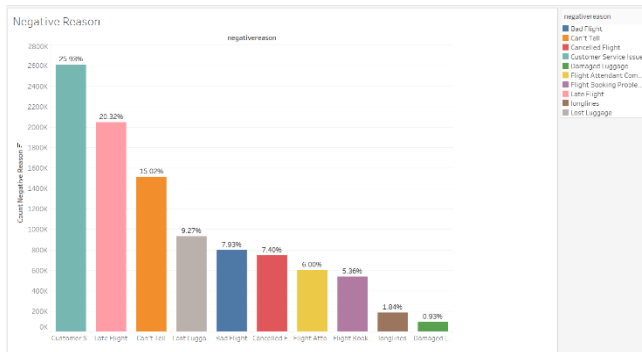


Fig. 15. Negative reason.

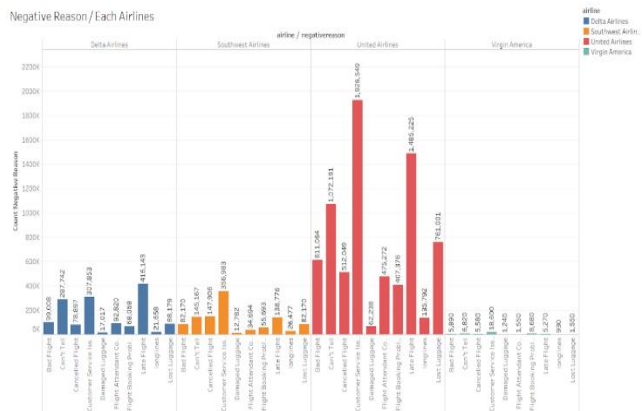


Fig. 16. Negative reason for each airline.

The United Airlines should understand why the customer was not satisfied with the services provided as it is the critical element to increase the number of customers. To achieve this, United Airline must prepare analysis with the negative reviews to see how they performed if compared with other airline companies, so that United Airline can know what they can improve to attract the customer. Fig. 16 above shows the dashboard which tells the United Airline negative reasons of the customer. The insights are not accurate for Virgin America because the amount of negative reasons for them are not sufficient to compare with other airlines. So, the insights will be focusing on the three other airlines which are Delta Airlines, Southwest Airlines and United Airlines. Based on the

result, we can know that the top three negative reasons are customer service issues, flight delay, and loss of luggage. Among them, most of the customers which are 1926549 feel unhappy because they were not satisfied with the customer service. Thus, the airline company needs to improve their customer service to the customers. For example, providing training to the cabin crew and the ground service crew which can greatly improve the customer experience. Besides, 1,485,225 customers feel unhappy because of the delayed flight. Flight delay may disrupt the customer's schedule, especially for the businessman. There are 761,001 unhappy customers because of their luggage lost during the transportation. It can be critical issues as it may cause customer losses on financial.

#### D. Airline Rating Analysis

The Fig. 17 shows the number of travelers with different traveler's types for different airlines. According to the figure, there are four traveler types which are business travelers, couple leisure travelers, family leisure travelers, and solo leisure travelers. The airlines include Delta Airlines, Southwest Airlines, United Airlines and Virgin America Airlines. It is obvious that the United Airlines has the most travelers regardless of the travelers' type. For the United Airlines, the count of each traveler type is different which have business travelers have the count 1750476, couple leisure travelers have 1819272, family leisure travelers have 1521156 and solo leisure traveler have 2721264. The fewest number of family leisure travelers travel on United Airlines. On the other hand, the number of solo leisure travelers traveling through United Airlines are the most. Meanwhile, from the figure, Virgin America gets the least travelers compared to other four airlines. This is because Virgin America is not the major airline in the U.S.

The Fig. 18 shows the traveler's average rating to the seat comfort between the different cabins. Usually, there are four types of cabins which are business class, economy class, first class, and premium economy. Different types of cabins, the price is also different, and the services provided are also not the same. According to the Fig. 18, the rating given by the passenger of the seat comfort is high and low, compared to the different airlines and different cabins. We can see that the highest average rating of seat comfort is given by the family leisure travelers of the business class of Southwest Airlines and the couple leisure travelers of the premium economy of Virgin American. Southwest Airlines and Virgin America Airlines have significantly different ratings of seat comfort given by passengers. In contrast, the average ratings given by passengers of Delta Airlines and United Airlines are not much different. Both get the 5 of average rating, this means that the travelers feel it is comfortable when they sit on the seat. On the other hand, the lowest average rating that we can view from the figure is given by the travelers, which are the couple leisure travelers of business class of Southwest Airlines and family leisure travelers of the premium economy of Southwest Airlines. The average rating is the 1 score given by travelers. The reason why travelers are given a low rating might be that they feel that the seat is not comfortable.

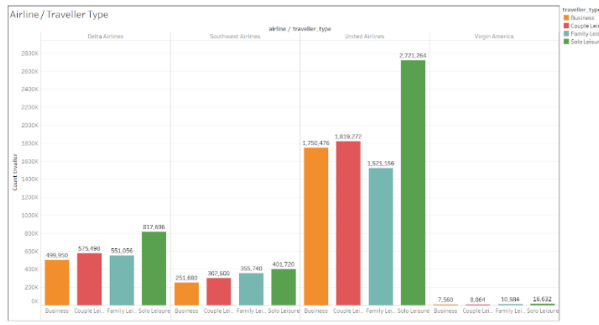


Fig. 17. Airline rating analysis dashboard.

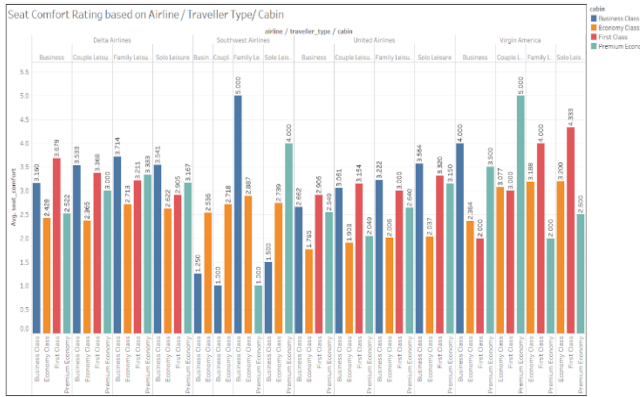


Fig. 18. Seat comfort rating.

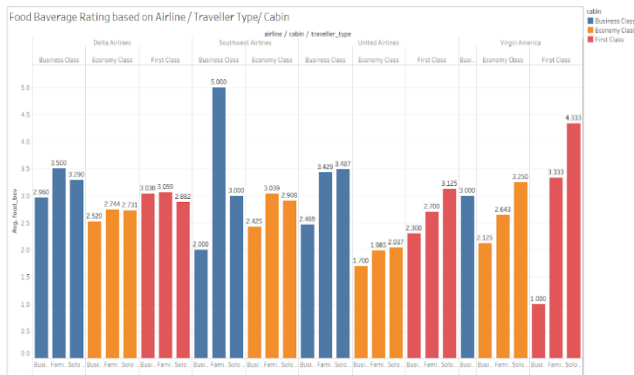


Fig. 19. Food beverage rating.

Usually, the stewardess on the plane will push the food truck to ask each passenger if they need food and beverage, or some passengers have already ordered the food and beverage they need when buying tickets. Therefore, the Fig. 19 shows how satisfied the passenger is with food and beverage. From the Fig. 19, passengers of different airlines have different levels of satisfaction with food and beverages. The passengers of Delta Airlines gave an average rating of satisfaction with food and beverage between 2.5 and 3.5. So, we can know that they are not very satisfied or dissatisfied with the food and beverages provided by the airline. We can see that the highest rating of food and beverage is given by the family leisure travellers of the business class of Southwest Airlines. Then, solo leisure travellers of the first-class cabin of Virgin America gave the second-highest rating for the food and beverage. The average rating for both is 5.0 and 4.3. This shows that passengers are very satisfied with the food and

drinks provided by the airline. On the contrary, the business travellers of the first-class cabin of Virgin America gave the lowest rating for the food and beverage. The second-lowest rating is given by the business travellers of the economy class of United Airlines. The average rating given by both is 1.0 and 1.7. The reason they are given low ratings may be that they feel that food and drink cannot satisfy them.

When passengers take a long-haul flight, they may feel bored on the flight. Therefore, some of the airlines have provided entertainment for their passengers to spend time. There are many entertainments on the plane for the passengers such as in-flight WI-FI, music library, TV shows and movies, games, digital shopping and so on. The Fig. 20 shows the average rating of passengers on the entertainment provided on plane. Overall, we can see that the first-class cabin of the passenger is more satisfied with the entertainments that provide on the plane than the passenger of the economy class. This is normal because the ticket price of the first-class cabin is higher than that of the economy class, so the entertainment provided is different. The highest average rating was given by the solo leisure travellers of the first-class cabin of Virgin America which have 4.5 of average rating. Because Virgin America has been voted as the airline which has the best in-flight entertainment for the passenger, that is why it got the highest rating. On the other hand, the lowest average rating from the above figure is business travellers of the economy class of United Airlines which have 1.725 of average rating. Meanwhile, the overall entertainment rating of the economy class and first-class cabin of United Airlines is lower than other economy class and first-class cabins of the airlines. Thus, United Airlines must figure out the problem and improve to provide better entertainment for its passengers.

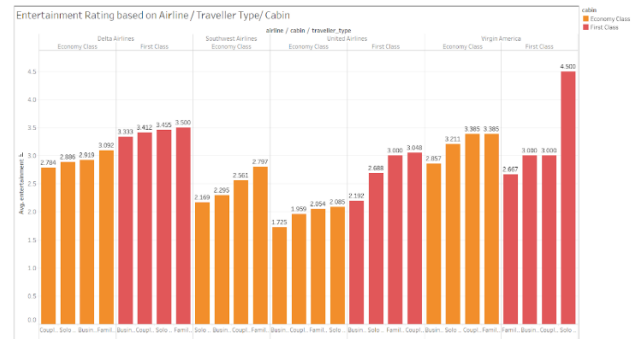


Fig. 20. Entertainment rating.

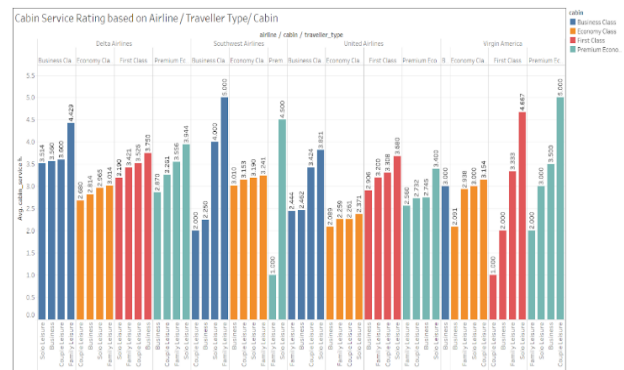


Fig. 21. Cabin service rating.



The Fig. 21 shows how satisfied the passenger was with the service provided by the stewardess in a different cabin. According to the graph, the rating given by the passenger is different. The cabin service provided by Delta Airlines overall is more average. Besides that, the average rating of Southwest Airlines and Virgin Airlines some are high, and some are low. The highest rating that was given by the family leisure travelers of business class of Southwest Airlines and couple leisure travelers of the premium economy class of Virgin America, both get 5 average ratings each. Meanwhile, the lowest rating that was given by the family leisure travelers of premium economy class of Southwest Airlines and couple leisure travelers of the first class of Virgin America, both get 1 average rating each. Thus, the airlines must figure out the problem and provide solutions to prevent loss of customers and achieve a good reputation. To prevent this, the airlines can provide training for their stewardess and other staff, instill the concept of customer first. In addition, give punishment for those who had bad service that was provided to the passenger.

## VI. CONCLUSION

In conclusion, this paper demonstrates the importance of business intelligence in the aviation industry. United Airlines needs to uncover hidden knowledge and gain a more competitive advantage. This study has analyzed the customer reviews on Twitter to understand customers' opinions and emotions on the aviation industry. The authors believe that the insights and BI solutions included in this study can help United Airline to improve their business values.

## REFERENCES

- [1] Taylor, D., 2020. *What is Business Intelligence? BI Definition, Meaning & Example*. [Online] Available at: <https://www.guru99.com/business-intelligence-definition-example.html> [Accessed 20 October 2020].
- [2] Tableau, n.d. *Business Intelligence vs. Business Analytics: What's The Difference?*. [Online] Available at: <https://www.tableau.com/learn/articles/business-intelligence/bi-business-analytics> [Accessed 20 October 2020].
- [3] Azure, M., n.d. *What are business intelligence (BI) tools?*. [Online] Available at: <https://azure.microsoft.com/en-gb/overview/what-are-business-intelligence-tools/> [Accessed 20 October 2020].
- [4] Guru99.com. 2020. *24 Best Business Intelligence(BI) Tools List In 2020*. [Online] Available at: <https://www.guru99.com/business-intelligence-tools.html> [Accessed 21 October 2020].
- [5] En.wikipedia.org. 2020. *United Airlines*. [Online] Available at: [https://en.wikipedia.org/wiki/United\\_Airlines](https://en.wikipedia.org/wiki/United_Airlines) [Accessed 21 October 2020].
- [6] Nalley, J., 2019. *The Goals and Objectives of British Airways*. [Online] Available at: <https://careertrend.com/air-taxi-pilot-description-19515.html> [Accessed 22 October 2020].
- [7] Market Realist. 2020. *Low-Entry Barriers Intensify Competition In Airline Industry*. [Online] Available at: <https://marketrealist.com/2014/12/low-entry-barriers-intensify-competition-airline-industry/> [Accessed 22 October 2020].
- [8] OpenJaw, 2019. *7 Challenges for Airlines*. [Online] Available at: <https://www.openjawtech.com/7-challenges-airlines-18/> [Accessed 22 October 2020].
- [9] Dallemand, J., 2020. *THE 5 TRUE COMPETITIVE ADVANTAGES OF DATA-DRIVEN AIRLINES*. [Online] Available at: <https://blog.datumize.com/the-5-true-competitive-advantages-of-data-driven-airlines> [Accessed 23 October 2020].
- [10] Kaggle.com. 2020. *Twitter US Airline Sentiment*. [Online] Available at: <https://www.kaggle.com/crowdflower/twitter-airline-sentiment> [Accessed 18 October 2020].
- [11] Datapine. 2020. *Data Warehousing And Business Intelligence: A BI Architecture Guide*. [Online] Available at: <https://www.datapine.com/blog/data-warehousing-and-business-intelligence-architecture/> [Accessed 29 October 2020].s
- [12] S, V., 2019. *Step by Step Guide to Dimensional Data Modeling*. [Online] Available at: <https://dwgeek.com/guide-dimensional-modeling.html> [Accessed 29 October 2020].

# Model Predictive Controlled Quasi Z Source Inverter Fed Induction Motor Drive System

D. Himabindu<sup>1</sup>, G. Sreenivasan<sup>2</sup>, R. Kiranmayi<sup>3</sup>  
Research Scholar, Dept of EEE, JNTUA, Ananthapur<sup>1</sup>  
Professor, Dept of EEE, PVKKIT, Ananthapur<sup>2</sup>  
Professor, Dept of EEE, JNTUA, Ananthapur<sup>3</sup>

**Abstract**—Ongoing advancements in inverters have offered pathway to high gain quasi Z source inverter Circuit (QZSIC). High gain QZSIC can be found between Semi Converter (SC) and three phase Induction Motor Loads (TPIML). This paper proposes suitable controller for closed loop controlled QZSIC-TPIML. This strive deals within improvement in time- response of QZSIC fed induction motor system. The objective of this effort tis to design a closed loop controlled QZSI\*fed-induction motor framework that provides a stable-rotor-speed. The QZSIC is settled to switch it to “3phase AC”. The yield of 3phase inverter is sieved before it’s applied to a‘3phase-Induction-motor’. Closed loop control of QZSIC-TPIML using SMC and MPC is simulated &their rejoinders are compared. The ‘Model Predictive controller (MPC)’ is acclaimed to retain persistent significance of-speed. The result obtained via MP-controlled QZS-IIMD-method is compared with Sliding mode-controlled (SMC) QZS-IIMD systems for change in input voltage. The wished-for MP controlled-QZS-IIMD method has benefits like fast settling-time and less steady state speed error.PIC16F84basedhardware for 0.5HP, QZSIC-IMDS is implemented.

**Keywords**—QZSIC; TPIML; CLSC; SMC; MPC; IMDS

LIST OF ABBREVIATIONS
QZSIC-Quasi Z source Inverter circuit
TPIML- Three Phase Induction Motor Load
MPC - ‘Model Predictive controller
SMC - Sliding Mode-Controller
IMDS – Induction Motor Drive System
CLSC- Closed Loop Semi Converter

## I. INTRODUCTION

‘The ZSI were having the solitary stage buck-boost amendment capacity’. "The framework was having an uncommon impedance circuit for coupling the basic circuit of converter with focal point for getting a particular brand name which wouldn't be acknowledged by using the common voltage source inverter and current-source-inverter".

An exchanged inductor semi Z-source inverter (SI-qZSI) showed a higher increase than a semi Z-source inverter (qZSI) while keeping nonstop information current. Like sustainable power age frameworks that outcome in low information voltage, SI-qZSI displays helping capacity that may not be sufficient at times. A voltage-lifting unit can be framed by exchanging one of the diodes in the switch inductor unit. A high move forward geography, for example, qZSI with

voltage-lifting unit (qZSI-VL), can be derived [1].

M. Padma priya, et al. [2] proposed a plan and reproduction of further developed diode helped voltage took care of three stage semi z-source inverter for photovoltaic application. Variable shoot through time span was produced by adding low recurrence voltage with steady voltage and the control signal for variable shoot was created utilizing support PWM control. The result nature of three stage QZSI was improved by supplanting the diode with dynamic power switch in Switch Boost Inverter (SBI) [2].

Another QZSI, which lessens input current waves, relative abundance of sounds created, and further developing the result voltage waveform, conveyed to the organization was presented. In the geography of the proposed converter, two batteries were put in lined up with capacitors in the circuit. The batteries will diminish current waves brought about by changes happened in light power and its bearing. Besides, a changed dynamic/receptive power control was introduced and applied in proposed QZSI (PQZSI) to control the genuine and responsive result force of the PV, separately [3].

This work was expected to upgrade both the powerful execution and effectiveness of SPIM for water siphon application at evaluated speed. Multi-objective streamlining was performed by utilizing the hereditary calculation utilizing a Maxwell-2D transient solver. The examination introduced the correlation between the streamlined and essential engine plans while thinking about different assembling limitations. The review proposed a structure for execution upgrade of SPIM while checking the ideal arrangement through responsiveness examination [4].

"Z-source T-type-inverter-for RES with PR" was recommended by Ozdemir. ‘Here in examination, an innovative Z-source T-type-inverter for network associated RES’ was proposed [5]. The proposed framework had capacity of heightening-voltage-level with no extra DC’DC-converter/transformer over the~Z-impedance position. The extent-of-the framework was decreased by disposing of DC’DC-converter / venture up-transformer prerequisite. Versatile-closed-loop-state-control framework-for-a3level-neutral-point-clamped-ZSI was given by Wolffe.

The mix of a novel modified QZS with a solitary stage proportioned ~hybrid 3level ~inverter so as to help the inverter 3level yield voltage was presented. The anticipated distinct stage MQZS-hybrid 3level ~inverter gave a sophisticated enhancement capacity & lessens the quantity of

~inductors in the –source-impedance, contrasted & equally the solitary stage 3 level three phase inverter QZSIC and the AC stage rectified to QZS 3phase inverter. AC stage rectified to QZSIC with decreased capacitance utilizing adjusted modulation and twofold recurrence ripple concealment control" was given by Li. Fathi recommended "Improved lift ~ZSI-with exchanged Z impedance".

The principle focal point of the examination was to explore which procedure mining calculation could prompt generation of procedure models that separate replay the occasions accurately with 100% degree of wellness, exactness, speculation and straightforwardness. The outcomes demonstrated that alpha calculation brought about the generalization of procedure models with great straightforwardness however with poor accuracy and speculation. Heuristic calculation brought about the generalization of procedure models with great exactness however with poor speculation and effortlessness. Model predictive calculation brought about generation of rather basic procedure models with great accuracy and speculation. Besides, the models/diagrams created through Model predictive-calculation could separate the majority of the cases accurately with 100% degree of wellness as an approval measure. Procedure mining and students' conduct investigation in a collective and electronic multi-table top condition was proposed by Porouhan.

S. Ozdemir deals with "Z-supply T-kind inverter for renewable electricity systems with proportional resonant controller [5]. M. Stempfle affords adaptive closed loop country manage machine for a 3 degree neutral factor clamped Z supply inverter [6]. A.V. Ho labored on Single section modified quasi Z supply cascaded hybrid five degree Inverter [7]. Reduction of common mode voltage and conducted EMI through three phase inverter topology is given by Ham [8]. E. Babari has advanced embedded switched Z source inverter [9]. Improving performance of OTT systems the use of fuzzy mining technique is given by way of Premchaiswadi [10].

V. Jagan, S. Das cope with two tapped inductor quasi impedance source inverter (2TL-QZSIC) for PV applications [11]. Sliding mode manage of the planar switched reluctance motor for interference suppression is given through Xu [12].

In this topic, the existing work is exposed with version predictive manage to improve the dynamic response of QZSI-IMDS and is organized in six sections. Section I presents the studies gap. In Section II, topology of QZSI-IMDS is given. The behaviour of QZSI-IMDS in closed loop is offered in phase 4. The manage strategies for QZSI-IMDS are evaluated in Section III and Section IV offers the experimental results for QZSI-IMDS. Conclusion is mentioned in the closing Section V.

There is a want to regulate the velocity of QZSI-IMDS because of exchange in load or supply voltage the usage of closed loop controller .Hence this paper in particular makes a specialty of identifying a premier controller for a QZSI-IMDS used for constant pace packages. Generally there is a more call for constant speed loads. A closed loop control is delivered inside the QZSI-circuit with a purpose to attain regular pace quickly.

The exceeding literature doesn't percent with OTT-filter for QZ-SIIMD method. This painting proposes OTT-clear out for QZSIC-IMDS. The exceeding literature doesn't percent with enhancement of small scale balance with the use of MP-managed-QZ-SIIMD-method. Hence, the present exertion offers with assessment of SMC and MP-managed-QZ-SIIMD-techniques. This exertion proposes MPC for QZSIC-IMDS.

## II. SYSTEM DESCRIPTION

DC is renewed to 3phaseAC by way of QZSI.OTT-filter out reduces harmonics. The response of OTTF is applied to TPIM. Speediness is delimited by way of SMC. The "block diagram of proposed SMC and MPC QZSIC system" is out in Fig. 1. Speediness is appraised with the mentioned speediness and the inaccuracy is targeted to a SMC/MPC. The "result of SMC/MPC" panels the PW of the rectifier. The function of speed loop is to regulate the speed of IMS.

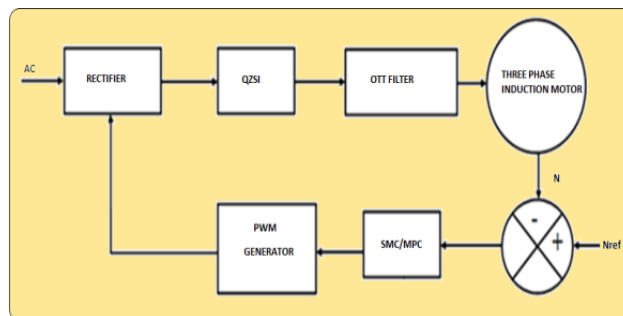


Fig. 1. 'Block diagram of SMC/ MPC controlled-QZSIC-system'.

## III. SIMULATION-RESULTS

Circuit diagram of CLSCQZSIIM with SMC is demarcated in Fig. 2. Supply Voltage is demarcated in Fig. 4. The-step-conversion is perceived and the assessment of supply Voltage is augmented from 200V to 230V at t =2.6 sec. Voltage across motor load of CLSCQZSIIM with SMC is demarcated in Fig. 5 and the value of Voltage across Motor Load of CLSCQZSICIM with SM controller is 340V. Current Through Motor Load of CL-SCQZSICIM with SM controller is demarcated in Fig. 6 and the value is 0.019A.

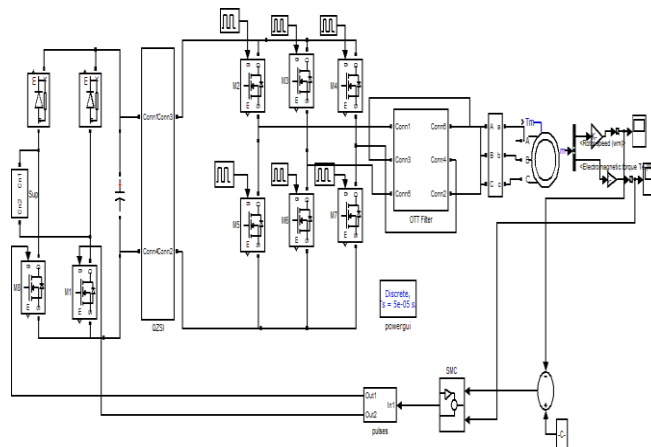


Fig. 2. CLSCQZSIIM with SMC.

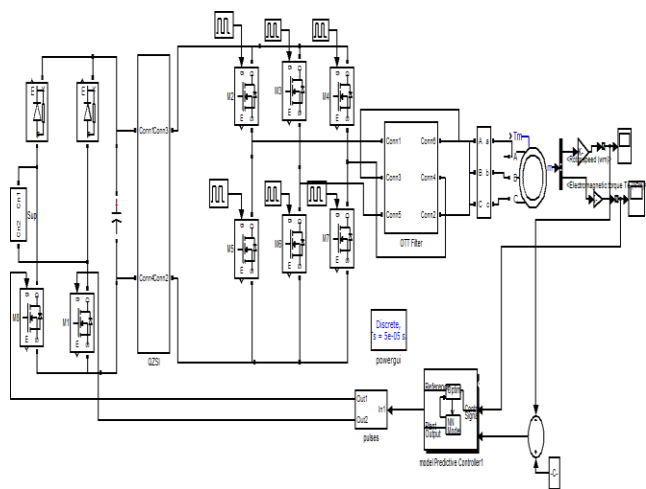


Fig. 3. CL-SC-QZSICIM with MPC.

CLSCQZSIIM with MPC is demarcated in Fig. 3. The MPC in the present work uses only speed and torque signals. Hence the CLSCQZSIIM with MPC is simpler. The contributions to MPC are speediness and torque.

Motor Speed of-CLSC-QZSICIM-with SMC and MPC is shown in Fig. 4. The significance of Motor Speed is 1296RPM. Speediness settles quickly with MPC. Number of speed-oscillations is higher with SMC than MPC.

Motor Torque of CLSCQZSIIM with SMC and MPC is shown in Fig. 5. The significance of Motor torque is 0.28N-m. Amplitude of torque oscillations is higher with SMC than MPC. Torque oscillations are lesser with MP-controller.

Evaluation of Time domain Parameters (speed) by means of SMC &MPC is specified in Table I. By means of MPC, the rise-time is reduced to 3.0Sec; peak-time is reduced to 3.18Sec; settling time is reduced to 3.36 Sec; steady-stateerror is reduced to 1.28RPM.

Evaluation of Time-domain-Parameters (motor-torque) by means of SMC &MPC is specified in Table II. By means of MPC, the rise-time is reduced to 3.12Sec; peak-time is reduced to 3.25 Sec; settling-time is reduced to 3.39Sec; steady-state-error is reduced to 0.03N-m. Hereafter, Closed-loop-QZSIC-fed-IM with MP controller is superior to Closed loop QZSIC-fed-IM with SMC.

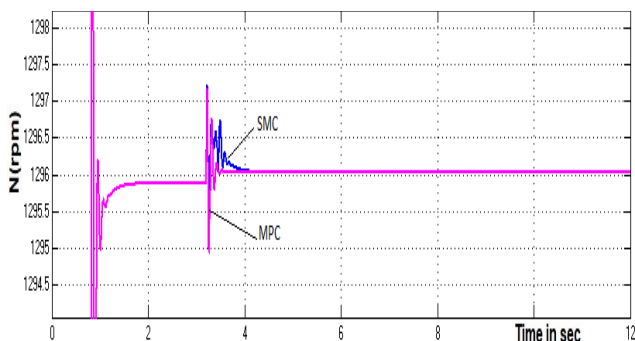


Fig. 4. Motor speed of-CLSCQZSIIM with SMC and MPC.

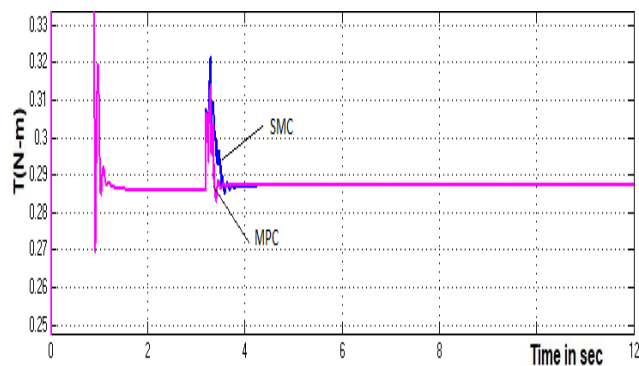


Fig. 5. Motor-Torque of CLSCQZSIIM-with SMC and MPC.

TABLE I. EVALUATION OF TIME DOMAIN PARAMETERS (MOTOR SPEED) BY MEANS OF SMC AND MPC

CL Controller	Rise-time (S)	Settling-time (S)	Peak-Time(S)	Steady-state-error (Rpm)
SMC	3.1	3.65	3.26	1.76
MPC	3.0	3.36	3.18	1.28

TABLE II. EVALUATION OF TIME-DOMAIN-PARAMETERS (TORQUE) BY MEANS OF SMC AND MPC

CL Controller	Rise-time (S)	Settling-time (S)	Peak Time(S)	Steady-state-error (N-m)
SMC	3.23	3.55	3.33	0.04
MPC	3.12	3.39	3.25	0.03

#### IV. EXPERIMENTAL-RESULTS AND DISCUSSIONS OF QZSICIMDS

QZSICIMDS hardware snap-shot is out in Fig. 6. The hardware embraces of rectifier, QZS, 3Ø-inverter, Transformer, Control-Circuit & 3Ø motor-load. PIC16F84 is used to generate pulses for the switches of QZSI. Current through motor load of QZSICIMDS is demarcated in Fig. 7. The spikes in current are due to change in switching sequence. Voltage of across motor of QZSICIMDS is shown in Fig. 8 and the spikes are due to change in current in inductance of IM.



Fig. 6. Hardware-snap-shot of QZSICIMDS.

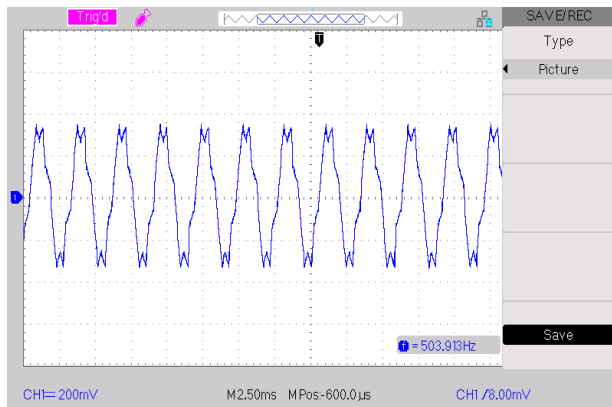


Fig. 7. Current through IMD.

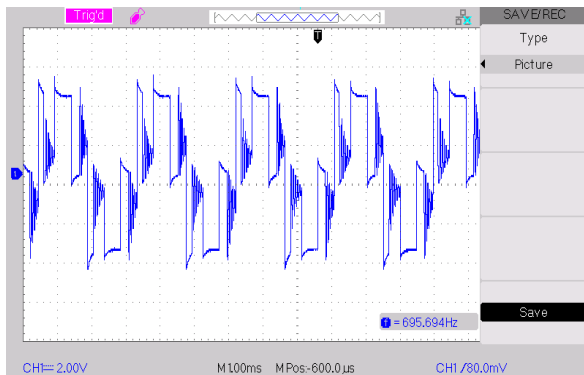


Fig. 8. Voltage across IMD.

## V. CONCLUSION OF SCQZSICIMD

The evaluation of SMC and MPC managed QZSIC is completed by way of Simulink. Simulation consequences are related via time domain parameters. SCQZSICIMD machine is efficiently intended, exhibited and-simulated the use of SMC and MP-Controllers for exchange in input voltage. By using MPC, settling-time is reduced to 3.39Sec; steady-state-error is reduced to 1.28RPM. The simulation consequences designate that, the response of MPCQZSIC is superior to SMC- QZSIC. The involvement of this exertion is to enrich the vibrant-reaction of QZSIIMD by MPC. The earnings of MPC-QZSIC are excessive expansion and fast dynamic reaction. Hence, first-rate voltage growth, amended performance and time retort of the machine makes the QZSIC much more proficiency making the SC-QZSICS suitable for input voltage assets. The downside is that SCQZSIC requires two additional controlled switches. The contribution of this work is to decorate small scale stability of QZSIIMD power using MPC and reduce THD using the use of OTT filter. This QZSIIMD is extra suitable for programs like electric motors and compressors.

The existent attempt pacts with “SMC & MP controlled QZSIC”. The evaluation among FLC & ANN QZSICS may be completed in-destiny. The hardware of QZSIIMD is applied using PIC16F84. The hardware of QZSIIMD can be implemented the usage of DSPIC to boom the switching frequency. High power QZSIIMD- may be carried out using IGBTs. Cost analysis of “SMC & MP controlled QZSIC” can be done in future.

## REFERENCES

- [1] Raheel Afzal Yu Tang, Yinghoo Song, “Comparative analysis of switched inductor based Quasi z source inverters” Springer, Journal of power electronics 2022, published 30 June 2022.
- [2] M. Padmapriya & T.A. Raghavendiran, Improved diode assisted voltage fed three phase quasi Z source inverter for photo voltaic application, Springer, Journal of ambient intelligence and humanized computing, 12, 5505-5512, 15 June 2022.
- [3] S. Honarbari & M. Alizadeh Bidgoli, Designing a Quasi Z source inverter with energy storage to improve grid power quality, <https://doi.org/10.1080/03772063.2019.1709571>, Taylor & Francis, IETE journal of research, published online 10 Jan 2020.
- [4] M. Bin Younas, H.A. Khalid, Adeel Javed, H. Yetis, T. Goktas & M. Arkan, Performance enhancement of single-phase induction motor using GA based multi-objective optimization, <https://doi.org/10.1080/00207217.2021.1969445>, Taylor & Francis, International journal of electronics, published online 05 Sept, 2021.
- [5] S. Ozdemir, “Z-source T-type inverter for renewable energy systems with proportional resonant controller”, International Journal of Hydrogen Energy, 2016.
- [6] M. Stempfle, S. Bintz, J. Wolfle, J. Roth Stielow, “Adaptive closed loop state control system for a three level neutral point clamped Z source inverter”, IET Electr. Syst. Transp., vol. 6, no. 1, pp. 12-19, 2016.
- [7] A. V. Ho, T. W. Chun, “Single phase modified quasi Z source cascaded hybrid five level inverter”, IEEE Trans. I. E., vol. 65, no. 6, pp. 5125-5134, Jun 2018.
- [8] C. T. Morris, D. Han, B. Sarlioglu, “Reduction of common mode voltage and conducted EMI through three phase inverter topology”, IEEE Trans on PE, vol. 32, no. 3, pp. 1320-1724, Mar 2017.
- [9] E. Babaei, E. S. Asl, M. H. Babayi, S. Laali, “Developed embedded switched Z source inverter”, IET PE., vol. 9, no. 9, pp. 1828-1841, Jul 2016.
- [10] P. Sirijaitham, P. Porouhan, P. Palangsantikul and W. Premchaiswadi, “Improving efficiency of OTT systems using fuzzy mining technique”, 2017-15th International Conf. on ICT and Knowledge Engineering (ICT&KE), Bangkok, 2017, pp. 1-5.
- [11] V. Jagan, S. Das, “Two tapped inductor quasi impedance source inverter (2TL-QZSIC) for PV applications”, Proc IEEE 6th Int Conf. Power Syst., pp. 1-6, Mar 2016.
- [12] Xu, Zhan-Zhi, et al (2016) Sliding mode control of the planar switched reluctance motor for interference suppression. Industrial Electronics and Applications (ICIEA), 11-Conference, 2016.

## AUTHORS’ PROFILE



D. Himabindu is now pursuing Ph.D. in the Department of EEE JNTUA University, Anantapur & also employed as an Assistant Professor Department of EEE in SOET, SPMVV, Tirupati AP India. Her-exploration interests comprise Power Electronics, Electrical Machines, & Modeling & Analysis of ACDrives. She is a life member of ‘ISTE’.



Dr. G. Sreenivasan employed as a ‘Professor Department of EEE’ in PVKKIT, Anantapur AP. His-extents of concern comprise Power Systems, Control Systems & Electrical Machines. He is a life member of “ISTE & IEI”.



Dr. R. Kiranmayi employed as a Professor in Department of EEE in JNTUA College of Engineering, Anantapur AP. Her extents of concern comprise Renewable energy resources & Electrical power systems. She is a ‘life member’ of “ISTE & IEI”.

# Visualization of Business Intelligence Insights into Aviation Accidents

## Aviation Accidents Discovery: Actionable Insights and Data Visualization

Loe Piin Piin<sup>1</sup>, Sarasvathi Nagalingham<sup>2</sup>

Faculty of Data Science and Information Technology, INTI International University  
Nilai, Negeri Sembilan, Malaysia<sup>1,2</sup>

**Abstract**—Despite the recent tragic loss activity, flying is often said to be the safest form of transport, and this is at least true in terms of fatalities per distance travelled. The Civil Aviation Authority reports that the death rate per billion kilometers travelled by aircraft is 0.003, which is much lower than the rates of 0.27 for train travel and 2.57 for vehicle travel. Despite the fact that safety has been the aviation industry's top focus for the last century and a half, accidents involving aircraft continue to be a source of horror even in the present day. Hence, the aim of this project is to identify the major causes and reasons that led to accidents in the aviation industry and to carry out research, finding, design, build and suggest a Business Intelligence (BI) solution to the problem. Throughout the project, it will discover problems, both elementary and critical which needs to be corrected or changed in order to prevent major negative happenings and improve the current situation in a positive way. Tableau will be the primary BI tool used in this process. Data visualization is the graphic depiction of information and data. Data visualization tools offer an easy approach to data analysis that observe and find patterns, outliers, and patterns in data by employing visual elements like charts, graphs, and maps. The project will also cover the initial to building and deployment stage of the BI solution to improve and prevent further accidents.

**Keywords**—Aviation; accidents; business intelligence; prediction; dashboard visualization; data analysis

### I. INTRODUCTION

#### A. Business Intelligence Methodology

Businesses can improve their decision-making processes with the help of Business Intelligence (BI), which showcases current and historical data within the context of the company's operations. BI may be used by analysts to establish the business's performance and competitive benchmarks, which will help businesses operate more seamlessly and productively and also uncover market trends that can be improved upon [14]. Every BI project has to specify a systematic methodology and strategy in order to provide decision-makers with the ability to adopt, implement, and integrate sustainable management practices across the company. In addition, it may improve the likelihood of accomplishment, reduce the amount of time and effort required, eliminate redundant procedures, and guarantee accurate reporting and analysis. As a result, the Polar's BI methodology will be used, which consist of five major steps. It helps organizations by delivering meaningful insights and develops new solutions to meet the problems based on

recommended steps. Fig. 1 depicts the stages that are included in the phase of constructing a Polar's BI methodology:

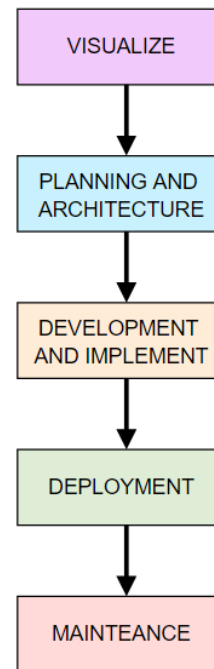


Fig. 1. Diagram for polar BI methodology

- **Visualize:** This step will involve defining the business requirement, which in our case, will be to identify trends and causes that leads to aviation accidents. It establishes a roadmap which will define the information clearly.
- **Planning and Architecture:** This step involves detailed business, data, and requirements by the system to produce the implementation plan accordingly including detailing the architecture of the system along with design specifications to provide solution. It is vital to define each requirement obtained from visualization and define a proper scope. It includes clearly defining the basic software and hardware requirements.
- **Development and Implement:** This phase is where the requirements are transformed into reality with a tested software production. Data will be populated into the

defined data structure through ETL process. After which database must be developed and software testing has to be conducted to ensure smooth functioning of the system and customer satisfaction. Finally, extensive documentation has to be prepared before deployment.

- **Deployment:** This is the implementation phase where the system will be finally implemented into the production environment to facilitate information exchange and migration of database structures and components. It gathers the readiness of the users and infrastructure, and the deployment plan will be shared with the initial set of users. The product would be deployed, and the system will be checked for correctness and accuracy.
- **Maintenance:** After the deployment, the system will need proper maintenance to ensure long term functionality and flow. This will include administration control, regular monitoring and performance updates and tuning.

### B. Overview of Boeing Company

Boeing Co., Major U.S. firm that is the world's biggest aviation organization and principal creator of business fly vehicles. It was established by William E. Boeing (1881-1956) in 1916 (as Aero Products Company). In the last part of the 1920s it turned out to be essential for United Aircraft and Transport Corp., however it reappeared as a free element in 1934 when that organization was separated to conform to antitrust regulation [1]. Boeing spearheaded the advancement of single-wing planes during the 1930s; its B-17 Flying Fortress (first flown 1935) and B-29 Superfortress (1942) assumed noticeable parts in World War II.

It is additionally a main maker of military airplane, helicopters, space vehicles, and rockets, a standing essentially improved with the organization's procurement of the aviation and safeguard units of Rockwell International Corporation in 1996 and its consolidation with McDonnell Douglas Corporation in 1997. Previously Boeing Airplane Company, the firm accepted that its present name in 1961 to mirror its venture into fields past airplane make. Central commands were in Seattle until 2001, when Boeing migrated to Chicago.

Boeing Company's constituent specialty units are coordinated around three primary gatherings of items and administrations business planes, military airplane and rockets, and space and correspondences. Boeing fabricates seven particular groups of business airplane, which are gathered in two offices Renton, and Everett-in Washington State and one office in California. The Renton plant fabricates the restricted body Boeing 737 and previously constructed the 757 airplanes (ended in 2004), while the wide-body Boeing 767 and 777 airplanes and a predetermined number of the plane, to a great extent suspended 747s is gathered at the Everett plant. The 787 airplanes are collected at the Everett plant and at an office in North Charleston, South Carolina. Boeing Business Jets, a joint endeavor of Boeing and General Electric Co., makes and markets business jets in view of the 737-700 aircraft as well as VIP forms of the 747, 777, and 787 carriers.

### C. Challenges Faced by Boeing Company

It is critical for businesses to be able to identify and forecast existing and future challenges that they may encounter. Identifying difficulties early on may assist the firm in developing a better solution sooner rather than later, which can help the company lower its losses. Aside from that, businesses may transform issues into chances for growth, allowing them to become more profitable in the long run. Being the largest producer of commercial and military aircraft and having the highest revenue compared to its competitors, the company has faced and faces several challenges.

First, Boeing being the leading company in the aviation sector has many responsibilities and must abide government-imposed norms without fail as well as manage their business operations without defect. Certain time sudden laws and restrictions can cause the company to change its business layout in order to adjust. For example, in 1977, Boeing faced travel cost competition when the US administration brought down air travel restrictions which promoted more newer airlines to offer the same services at very competitive prices by the year 1999. Another scenario of rapid change was during the Covid pandemic which affected many businesses but demanded a complete shutdown of aviation travel.

As the making of an airplane is directly related to safety, it has to undergo various tests in order to prove its feasibility and safety mechanisms. These include various additional costs and takes a long time. When Boeing developed the '787 Dreamliner' to save fuel and minimize stopovers, the airplane was made to be 50% carbon fiber-reinforced plastics, 12% titanium and almost 80% reduction of the use of aluminum which was primarily used in the old plane models. When it decided to use Lithium-ion batteries there was a concern that it would overheat to a level where it would stimulate fire. In order to solve this issue, it had to undergo various tests to determine its feasibility.

Besides, there is a high risk of the parts not being suitable and face trouble when interlocking since the company has partnerships and contracts with many other production companies for different parts of the plane. Therefore, they have to be returned back in order to be reproduced which increases the initial cost and also is very time consuming. Also, the aviation industry is very much dependable on service and the satisfaction of passengers. Therefore, the airline production companies have to meet the demands of the purchasing airlines in order to produce the model that will meet current consumer expectations as well as balance business costs.

Technological advancement is one of the major challenges of the industry. In order to implement newer software, sometimes with additional measures, the construction plan of the flight model has to be altered in order to compile it with the newer changes. These may take longer times as it must be planned and the whole company has to be suited with the model in order to train the employees such as engineers, software developers and other employees. Changes like this cannot be executed frequently as it is a very complex process, and it will also require additional training for existing pilots all around the world. Suppose the technical advancement is not

considered to be great, airline will switch companies to buy airplanes in order to save their pilots from additional training which takes up time as well as money.

#### D. Opportunities of Boeing

Boeing has always been at the top among aerospace-based organisations. However, Boeing still faces some disparity with other modern gaming companies. Based on the studies found on the current environment of the company and gaming industry, there are some key business opportunities that can be implemented into the standard operation procedure of the company.

The first opportunity that presents itself to the organisation is in the field of technology application. For example, the sharing of digital information among air traffic control, the flight deck, and the operations centre of an airline to enhance the effectiveness of flight routing and ensure passenger safety. An electronic flight bag software that utilises next-generation communications may provide information on alternative flight paths to pilots in the event that the conditions of the weather so need. In addition, connected cabin technologies that enhance common areas, such as galleys and restrooms, as well as monitor environmental factors, like as temperature and humidity allow automatic adjustments. Lastly, the installation of cameras will provide a greater number of passengers a view of the world beyond the window of the aeroplane.

Over the last several decades, safer and more dependable designs have been responsible for most of the progress gained in lowering accident rates and boosting efficiency. This breakthrough has been made possible by advancements in engines, systems, and structures. Furthermore, design has long been acknowledged as a component in minimizing and managing human mistake. When Boeing embarks on a new design activity, previous operating experience, operational objectives, and scientific understanding establish the human factors design criteria. To examine how well alternative design solutions fit these objectives, analytical approaches such as mockup or simulator assessments are performed. A human-centered design philosophy underpins this work, which has been confirmed by millions of flights and decades of experience.

When there is a big change in the design of a plane it has to undergo various tests which consume time and large sum of money. Since bringing lengthy changes to a plane is necessarily not a desired output in the short term, the company can adjust only the required mechanisms and not change the whole flight design which further causes workers to consume more time and gain used to the working process. This can be achieved by analyzing the areas which can be improved only to the level where it doesn't require additional guarantee tests and slowly bring changes over time to maintain balance in the business operations and revenue as well as providing a safe machine to its consumers.

Finally, the 787 can travel farther distances than its predecessors because to its improved fuel efficiency. As a result, it has enabled the creation of more than 50 nonstop routes throughout the globe. The design and construction of the Dreamliner are both ground-breaking. New composite

materials are used in the construction of fifty percent of the principal structure of the Boeing 787, which includes the fuselage and the wings. Both the speed of the aircraft and its fuel efficiency may be boosted thanks to the design and construction of the wings.

## II. PROPOSING BI SOLUTIONS

Business intelligence is a term that encapsulates the processes and methods used to collect, store, and analyze data from business operations or activities in order to improve performance [10]. All of these factors combine to provide a full image of a company, allowing customers to make better, more decisive decisions. In this section, the proposed business intelligence solutions assisting the business opportunities and solving the problems would be discussed.

Accident or crash rate is a calculation between the number of aircraft crash that unfold in a given period (crash frequency) and number of flights conducted within the same period. BI tools enable the accident rate to be calculated faster. In addition, the metrics for calculating the crash rate can be analyzed individually too to provide more insights, such as the crash frequency. Thus, the company can predict the probability of crash happening in the following years and put up a countermeasure to reduce it.

Through BI tools, the company will be able to discover factors that contribute to airplanes crashes. It allows the company to correlate the cause of the accidents, either its human errors or machine malfunctions, and any other piece of information together. Based on the piece of information founded, the factors that lead to aircraft crash can be grouped based on certain criteria to provide better insight which can also be used for comparison purposes. Therefore, a more specific study of aircraft crash on each criterion can be performed and a plan to address each category problem can be addressed.

Predictive maintenance is a proactive approach to machine maintenance, which is made possible through utilizing BI tools. It provides the company to organize schedules based on continuous condition monitoring. To minimize additional costly failures, broken parts are fixed or replaced if unfavorable patterns are discovered. Lower maintenance costs, longer equipment life, less downtime, increased production capacity, and improved safety are just a few of the advantages the company can expect.

## III. DATASET ANALYSIS

To describe how BI can improve Boeing performance, the NTSB aviation accident dataset was used as our data source. NTSB is an independent U.S. government investigation agency that identify and report on aviation accident. The NTSB aviation accident database contains information on civil aviation accidents and selected incidents that occurred in the United States, its territories and possessions, and international waterways. For our research, the dataset that we use contains information on aviation accidents from 1962 to present [5].

The reason the records are up to date is made possible by the continuous update of the accidents into the database. First,



a preliminary report will be available online within a few days of an accident. When available, information is added, and when the investigation is complete, the preliminary report is replaced with a final description of the accident and its probable cause. Full narrative descriptions may not be available for cases prior to 1993, those under revision, or those in which the NTSB did not have primary investigative responsibility. It's worth noting that the information isn't limited to just commercial jets. On September 18, 2002, data from 1962 to 1982 were added to the aircraft accident database. The structure and type of data given in previous briefs may change from that contained in subsequent reports.

An aviation accident is described as an event occurring due to aircraft operation happening between the time people boards the aircraft with the aim of flight and the time people depart, in which people are mortally or severely injured, the aircraft sustains severe damage or structural failure, or the aircraft goes missing or becomes totally inaccessible, according to Annex 13 of the Convention on International Civil Aviation. An aviation incident is defined in Annex 13 as any occurrence, other than an accident, related with the operation of an aircraft that affects or has the potential to influence the safety of operation. Government agencies such as the FAA and the NTSB examine accidents and occurrences.

A. Description of Dataset

Our dataset is obtained NTSB from aviation accident database containing information about civil aviation accidents within the United States and its territories including international water from the year 1962. The dataset comes in a single data sheet and consists of 31 columns describing details about the recorded aviation accident [5]. The table below shows the detailed description of each attribute in the dataset and what it represents (Table I).

TABLE I. DESCRIPTION OF DATASET

No	Attribute Name	Data Type	Description
1	Event ID	ID	Unique alphanumeric for each of the row about the accident
2	Investigation Type	ID	Conveys if the scenario was an accident or incident
3	Accident Number	ID	Specific number given to the accident/incident
4	Event Date	Date	Date at which the accident occurred, displayed in the format "YYYY-MM-DD"
5	Location	String	Location at which the accident took place
6	Country	String	Country the accident took place
7	Latitude	Numeric (Continuous)	Angle of latitude which ranges from 0 degree at the Equator to 90 degrees (North or South) at the poles
8	Longitude	Numeric (Continuous)	Longitude which is the measurement east or west of the prime meridian
9	Airport Code	String	The airport code, which is unique for each and every airport in the world and is usually represented by three letters or four letters

10	Airport Name	String	The name of the Airport
11	Injury Severity	Nominal	Severity of the injury
12	Aircraft Damage	Nominal	The damage level to the aircraft
13	Aircraft Category	Nominal	Type of aircraft, ranging from Airplane, helicopter, glider, etc.
14	Registration Number	ID	Registration number of the aircraft
15	Make	Nominal	Brand type of the aircraft
16	Model	Nominal	Model number of the aircraft
17	Amateur Built	Boolean	True or false statement confirming if the aircraft is amateur built, meaning if they were built by people for their own education and recreation and not officially by an enterprise
18	Number of Engines	Numeric (Discrete)	Number of engines the aircraft had
19	Engine Type	Nominal	Type of engine the aircraft had
20	FAR Description	String	Description by Federal Aviation Regulations
21	Schedule	String	Represents the aircraft's schedule
22	Purpose of Flight	Nominal	Purpose of the flight, whether personal, instructional or others
23	Air Carrier	String	Airline company of the aircraft
24	Total Fatal Injuries	Numeric (Discrete)	Total number of fatal injuries due to the accident
25	Total Serious Injuries	Numeric (Discrete)	Total serious injuries due to the accident
26	Total Minor Injuries	Numeric (Discrete)	Total minor injuries due to the accident
27	Total Uninjured	Numeric (Discrete)	Total number of people uninjured after the accident
28	Weather Condition	Nominal	Condition of the weather at the time of accident
29	Broad Phase of Flight	Nominal	Phase of flight when the accident occurred
30	Report Status	Nominal	Status of the accident when reported
31	Publication Date	Date	Publication date of the accident

B. Data Preprocessing

With the dataset obtained from NTSB, we went through several rounds of data cleaning and preparation to make sure the data is in a clean format and is ready to be used for our BI dashboards. To perform data cleaning and preparation, we have used two programs, namely Microsoft Excel and R Studio. The programs were used to look at the state of the dataset (i.e. number of missing values for every column). We will go further and explain the techniques that we used in these two programs to aid our data cleaning and preparation process [6].

Microsoft Excel is an excellent software owned by Microsoft incorporation. It serves the purpose to organize, clean, arrange and simplify the data. Among the hundreds of features it has, some are being mentioned below.

1) *Conditions based formatting*: If you want to format the data based on conditions, then MS Excel is the best tool. For instance, you want to see only the Boeing Planes with a

seating capacity of less than 300 people. Then you can apply these conditions and see them. To use this feature, you need to select the desired column. In this case, the Fig. 2 demonstrates Column "Seating Capacity" is selected:

	A
1	Seating Capacity
2	500
3	400
4	325
5	235
6	215
7	150
8	200
9	575
10	600

Fig. 2. Select "Seating Capacity" column

Once you have selected the column go to "Home" option in toolbar and click on "Conditional Formatting" as shown in Fig. 3.

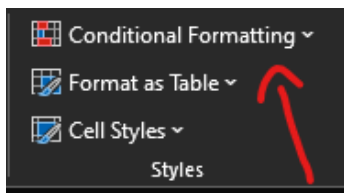


Fig. 3. Selection of "Conditional Formatting"

Once you hover the mouse over it, you can see the following options as shown in Fig. 4.

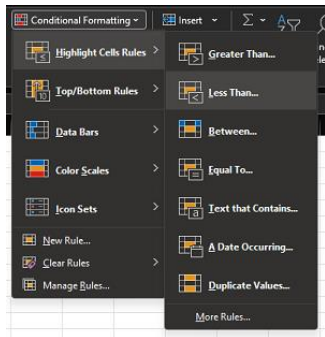


Fig. 4. Options in "Conditional Formatting"

In this case I have selected "Less Than..." as a condition. Once we click on "less than..." as shown in Fig. 5.

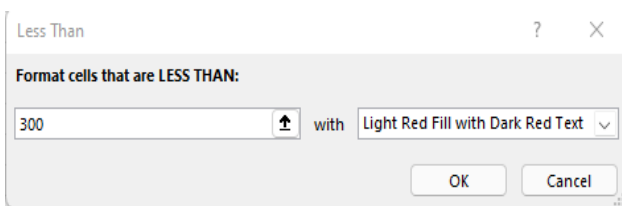


Fig. 5. Options in "Less Than..."

Here we can put the desired value and then the results will highlight in the column as depicted in Fig. 6.

	A
1	Seating Capacity
2	500
3	400
4	325
5	235
6	215
7	150
8	200
9	575
10	600

Fig. 6. Results of "Conditional Formatting"

2) *Removing duplicate columns:* There is a high probability that a data might have duplicate columns. MS excel provides a fantastic feature to remove them. You don't have to select each column in Dataset to remove them. Instead, there is an amazing feature in Excel which lets you remove duplicate columns automatically as shown in Fig. 7.

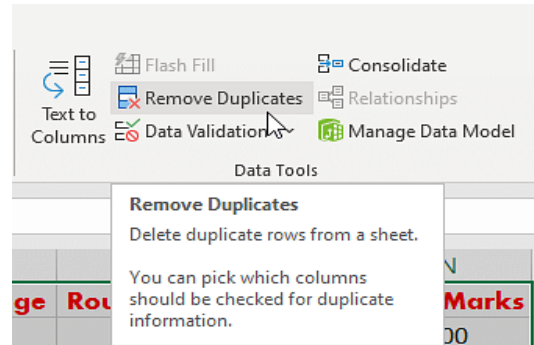


Fig. 7. Remove duplicates feature in excel

To access this feature, you can go to Data option placed in tool bar, there you can see "Remove Duplicate" feature as Fig. 8. Once you click on it, it will show you the duplicates and you can remove them if you want.

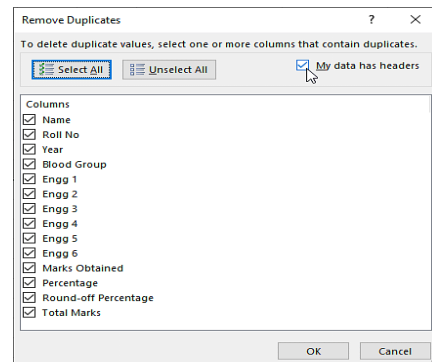


Fig. 8. Remove duplicates window

Once you are satisfied with columns checking you can hit ok and excel will remove duplicate column successfully.

3) *Spell check*: MS Excel has another useful feature as shown in Fig. 9 which helps you check for spelling mistakes in a data. This feature can come very handy in pointing out typos. For this feature to work, you can select your desired column from database and then click on “Review” option from tool bar. There you can see spell check option.

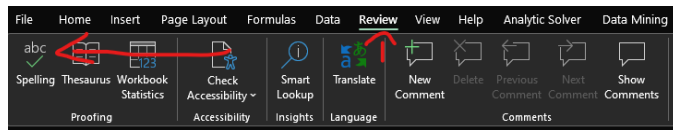


Fig. 9. Spelling check in excel

Once you click on it, Excel will automatically scan for any spelling mistakes in the data and will show as per the Fig. 10.

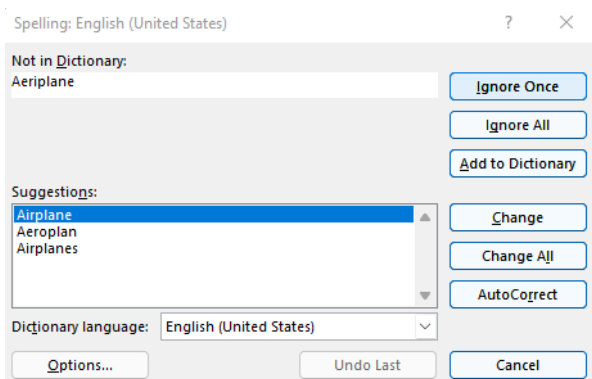


Fig. 10. Spelling check configuration

You can see this Pop-up, from here you can choose which spelling is suitable for you. If you think the word spell is ok, you can click on “Add to Dictionary” or “Ignore All/once”. If you want to change the word with the given suggestions, you can highlight preferred suggestion and click on “Change /Change All”. If you want the system to automatically correct, you can click on “Auto Correct”. If there are no spelling errors the following pop up will appear as shown in Fig. 11.

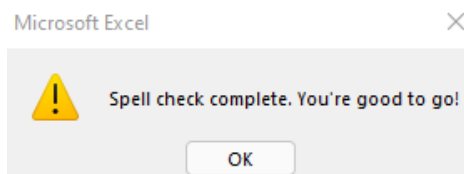


Fig. 11. Successful configuration of spelling check

RStudio is an open-source IDE (integrated development environment) for R programming that has been widely utilised by the community for statistical and computing purposes. The RStudio was utilised in this project since various tasks required the use of the R language to aid in data visualisation and pre-processing. Due to the ease of access and pre-experiences of the program being used, the authors decided to use this program as the pre-processing tool due to the familiarity of library tools usage from the past and the

flexibility of data processing it provides. Fig. 12 demonstrates the techniques that were used in R Studio to perform data cleaning and preparation:

4) *Display summary of a data frame*: `summary(df)`

```
> summary(avia)
  Event.Id      Investigation.Type Accident.Number
Length:87282  Length:87282      Length:87282
Class :character  Class :character  Class :character
Mode :character   Mode :character   Mode :character

  Airport.Code      Airport.Name      Injury.Severity
Length:87282      Length:87282      Length:87282
Class :character  Class :character  Class :character
Mode :character   Mode :character   Mode :character
```

Fig. 12. Display summary of non-numerical attributes of a data frame

Using this function, we will get a summary of every attribute in a data frame as shown in Fig. 12. For each attribute, it will display the number of records and the attribute type.

```
Total.Serious.Injuries Total.Minor.Injuries Total.Uninjured
Min. : 0.000      Min. : 0.000      Min. : 0.000
1st Qu.: 0.000    1st Qu.: 0.000    1st Qu.: 0.000
Median : 0.000    Median : 0.000    Median : 1.000
Mean : 0.281      Mean : 0.361      Mean : 5.312
3rd Qu.: 0.000    3rd Qu.: 0.000    3rd Qu.: 2.000
Max. :161.000     Max. :380.000     Max. :699.000
NA's :12510       NA's :11933       NA's :5912
```

Fig. 13. Display summary of numerical attributes of a data frame

For numerical attributes, R will be able to display even more details of the attribute, like the first quarter value, third quarter value, minimum and maximum value found, number of missing values etc as shown in Fig. 13. Using this function, we will be able to check the type of every attribute and also the number of missing values. However, as this function is unable to display more information about non-numerical attributes, we have to use other functions to learn more about our dataset.

5) *Display number of missing values or empty columns for each attribute*:

```
colSums(is.na(df)| df == "")
```

```
> colSums(is.na(avia)| avia == "")
  Event.Id      Investigation.Type      Accident.Number      Event.Date
0            0                0                0
Longitude      Airport.Code      Airport.Name      Injury.Severity
54218          37812            35254            883
Make           Model            Amateur.Built      Number.of.Engines
62            89                102            5749
Purpose.of.flight Air.carrier      Total.Fatal.Injuries Total.Serious.Injuries
5789          71311            11401            12510
Broad.phase.of.flight Report.Status      Publication.Date
25558         5211             13653
```

Fig. 14. Display number of missing values or empty columns for each attribute

Through this function, we are able to clearly see the number of missing values and empty columns every attribute has as shown in Fig. 14. Any missing values or empty columns are bad for making visualizations through BI dashboards. Thus, we have to figure out what to deal with attributes that have an overwhelming number of missing values or empty columns.

6) *Remove columns from a data frame*:

```
df = subset(df, select = -c(attribute1, attribute2, attribute3))
```

With this function, we are subsetting attributes to create a new data frame. However, as we are specifying the “-c” argument, we are actually creating a new data frame that excludes the attributes mentioned in the command. We used this function to remove any attributes that have way too many missing values or empty columns to be used in our BI dashboard visualizations.

7) *Remove rows with missing values:*

```
df <- na.omit(df)
```

We used this function to remove the rows that had only missing values in our data frame.

8) *Format values in an attribute:* In our data set, the attribute that mentions the injury severity of aviation accidents also mention the number of fatal injuries in Fig. 15 (i.e. Fatal(4) meaning four meaning suffered fatal injuries). As the data set already has an attribute that stores the number of fatal injuries, we can make this attribute to only mention whether the injury severity is fatal, non-fatal or incidental. To achieve this, we used two functions:

```
unique(df[c("attribute1")])  
  
> unique(avia_clean[c("Injury.Severity")])  
      Injury.Severity  
1          Fatal(2)  
2          Fatal(4)  
3          Fatal(3)  
5          Fatal(1)  
6        Non-Fatal  
24         Incident  
26          Fatal(8)  
85          Fatal(78)  
166         Fatal(7)  
436         Fatal(6)  
608         Fatal(5)  
1872        Fatal(153)
```

Fig. 15. Display unique values in an attribute

Fig. 15 shows us the unique values in an attribute. Using this command, we first checked whether the attribute had a big number of unique attributes, which is bad for a categorical attribute.

```
df$attribute1 <- sub("Searched sting", "Substitute string",  
df$attribute1, ignore.case = TRUE)
```

```
avia_clean$Injury.Severity <- sub("Fatal\\(\\d+)", "Fatal", avia_clean$Injury.Severity, ignore.case = TRUE)
```

Fig. 16. Command to replace strings in an attribute

As shown in Fig. 16, we managed to automatically clear out the number of fatal injuries in this attribute, leaving only the word “Fatal”. Displaying the unique values again as per the Fig. 17, the command was successful in clearing out the number of fatal injuries in the attribute.

```
> unique(avia_clean[c("Injury.Severity")])  
      Injury.Severity  
1          Fatal  
6        Non-Fatal  
24         Incident
```

Fig. 17. Display unique values after replacing strings

9) *Read and write files:*

```
df <- read.csv("File location path")
```

```
#Load csv file  
avia <- read.csv("C:\\Users\\Daniel\\Documents\\Personal Files\\School Files\\Unif\\IBM4207 Business Intelligence  
\\Group Assignment\\archive\\aviationdata.csv")
```

Fig. 18. Command to read CSV file into a data frame

Fig. 18 depicts that we could read the csv format file that our data set came in. A csv file is a delimited text file which uses commas to separate between different values, and using this function, we imported the file into R Studio as a data frame to perform data cleaning and preparation.

```
write.xlsx(df, "File location path")
```

```
#write into excel document  
write.xlsx(avia_clean, "C:\\Users\\Daniel\\Documents\\Personal Files\\School Files\\Unif\\IBM4207 Business Intelligence  
\\Group Assignment\\aviationclean.xlsx")
```

Fig. 19. Command to write data frame into an excel document

After our data cleaning and preparation is completed, we want to export our data frame into an Excel file so that it can be used for our BI dashboards. We used this function, which is part of the “openxlsx” R library to write our clean data frame into an Excel file as shown in Fig. 19.

Due to challenges during data recording and such, several attributes of our dataset had a large number of missing values/empty columns. For example, the air carrier column had 71,311 missing values, longitude and latitude had 54,218 and 54,209 missing values each. As our dataset has 87,282 records in total, including attributes like these with a large number of missing values will be problematic. We also cannot try to fill in the missing values with techniques like using the mean as it will distort the meaning of the data. Because of this issue, we decided to outright remove the attributes that has a large number of missing values.

C. *Data Modeling*

The process of generating a visual representation of a comprehensive information system or specific components of that system in order to express the relationships that exist between different data points and organizational structures is known as data modelling [4]. The purpose of data modelling is to offer illustrations of the many kinds of data that are utilized and stored inside the system, as well as representations of their connections, potential groupings and organizational structures, format types, and attribute values.

In data mining, an algorithm is a set of related computations and heuristics that are used to construct a model based on a collection of data. Before designing a model, the algorithm does an initial analysis on the data that people provide, looking for different sorts of patterns or trends. The results of this study are used by the algorithm, which goes through a number of rounds to determine which parameters will provide the most accurate mining model [12]. After that, the whole of the data gathering is analyzed using these criteria in order to derive actionable patterns and comprehensive statistics. In this section, the algorithms that were employed are the decision tree method, the Support Vector Machine (SVM) algorithm, and the Naïve Bayes algorithm.

1) *Naïve bayes algorithm*: The Naïve Bayes algorithm is a classification approach that is based on the Bayes Theorem and operates on the assumption that predictors are independent of one another [3]. To put it simply, a Naïve Bayes classifier is one that operates on the assumption that the existence of one feature in a class has absolutely no bearing on the presence of any other feature. The Fig. 20 illustrates an example of the formula for Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Fig. 20. Calculation for bayes' theorem

- The possibility that a certain hypothesis (A) will in fact come to pass is quantified by the posterior probability, abbreviated as P(A|B).
- Likelihood Probability, abbreviated as P(B|A), is a measurement that determines how probable it is, based on the data that is now available, that a particular hypothesis is right.
- Priority probability, sometimes abbreviated as P(A), refers to the possibility that a hypothesis exists prior to seeing the data.
- The likelihood of evidence is referred to as P(B), which stands for marginal probability.

2) *Decision tree*: A decision tree is a technique of supervised learning that may be used with both discrete and continuous variables in its analysis [6]. The most essential aspect of the dataset is used to determine the subgroups that are created from the dataset. The algorithms are what decide how this characteristic is categorized by the decision tree and how the divisions are made between the categories. The terminal or leaf nodes, which do not further split, are created when the root node, which represents the most significant predictor, divides into decision nodes, which are sub-nodes as shown in Fig. 21 [7].

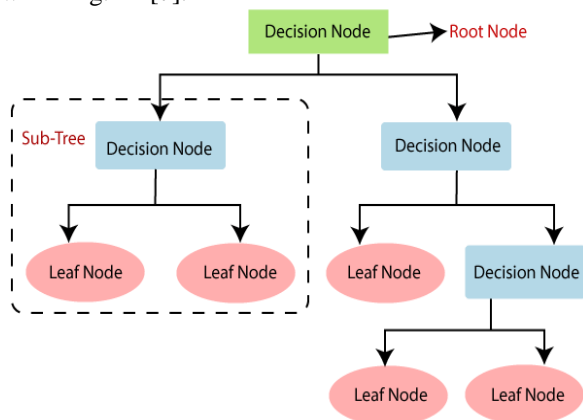


Fig. 21. Decision tree structure

3) *Support vector machine (SVM)*: The objective of the Support Vector Machine technique is to locate, in a space of N dimensions (where N represents the number of features), a hyperplane that categorizes the data points in a way that is unambiguous [8]. The two distinct groups of data points may be partitioned using any one of several hyperplanes that are available to choose from [9]. Finding a plane that has the highest margin, or the greatest separation between the data points of both classes, is the goal here. As per Fig. 22, When the margin distance is increased to its maximum, more support is added, which in turn increases the level of confidence with which subsequent data points may be categorized.

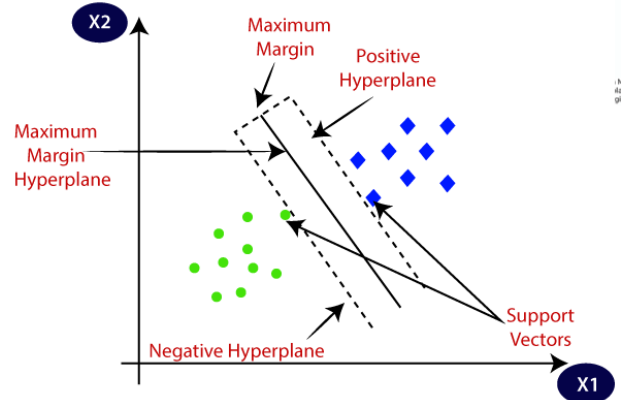


Fig. 22. SVM algorithm

#### IV. BUSINESS INTELLIGENCE ARCHITECTURE

The framework that a business establishes in order to run applications for business intelligence and analytics is referred to as a business intelligence architecture. It discusses the information technology platforms and software tools that are used to collect, integrate, store, and analyze business intelligence data before providing it to high-level executives and other corporate users as information on operational processes and statistics. Implementing an efficient BI program that allows data analysis and reporting to assist an organisation in monitoring business performance, optimizing business processes, recognizing new revenue opportunities, working on improving strategic planning, and providing better decisions overall requires a critical component known as the underlying BI architecture.

The BI architecture design shows how all the data and processes are connected to each other to support Valve business. The diagram shows the seven main components and the flow in Valve when dealing with data through business intelligence [2]. The seven main components consist of data sources, ETL (extract, transform and loading), data repository, business intelligence tools, interfaces, data management and security management as shown in Fig. 23.

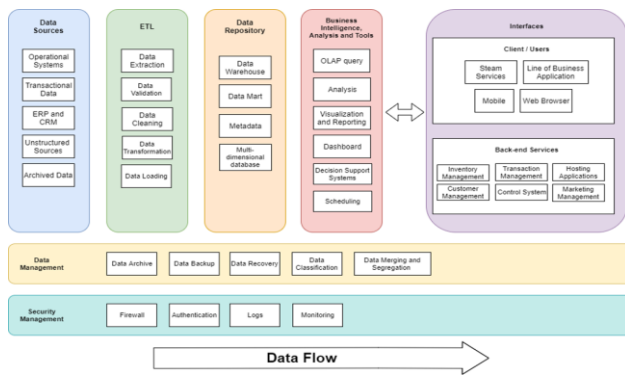


Fig. 23. Proposed BI architecture

### A. Data Source

Data sources mainly refer to where data is obtained. Any business organisation would get data differently depending on their mode of business, and their operations done. In Valve, data can come from multiple sources. For instance, transactions or purchases on games and the accessories can be categorised into transactional data, whereas daily processes and routines would contribute to data obtained from operational systems. This architecture design suggests that data sources may also involve data from unstructured forms, archives, and other systems such as ERP and CRM. All these data will then be staged into the ETL stage.

### B. ETL (Extract, Transform and Load)

This stage involves three major steps in dealing with data. First, data identified from the various data sources are extracted. There are multiple methods to extract data. It is crucial to discover which extraction method and platform would benefit the business the most. Usually when data are extracted from various data sources, the data vary from each other in terms of formatting. For instance, two data sources might store “date” data differently from each other. This is where the second step, transformation comes in. Transformation relates to changing the data into the wanted and appropriate format. This step also involves data cleaning and attributes derivations. This step is significant to ensure data integrity and quality, while making sure that the data would reach a common dimension and storage method. The data will then be loaded into the repository once all the data has been transformed.

### C. Data Repository

Data repository refers to where data is stored after being processed. Valve stores over ten million of players’ data, and management side. Therefore, Valve will need numerous data repositories and in different forms, such as data marts and data warehouses. For example, data that are generated to players and require high availability would be stored in data marts, and data that are not frequently used can be placed into data warehouses. Valve can store summarised and transformed data in data marts for respective departments to ensure productivity, while metadata that describes data and the relationships, together with raw data can be stored in a data warehouse. Multidimensional databases can help integrate data warehouses and OLAP applications to generate complex data to users [11].

### D. Business Intelligence, Analytics and Tools

This section comprises of all the business intelligence tools used in Valve. All of these tools assist the organisation in making better decisions and supporting the interfaces on both clients and back-end services. The tools include OLAP operations, analysis, visualisation, reporting, dashboard, and decision support applications.

### E. Interfaces

The interfaces involve how input and output would result through different platforms such as mobile, web applications, line of business application and steam application. Interfaces comprise of back-end services from within the company and externally from third-party services. These services extend to aid in the management side in Valve, such as inventory management, transaction management, customer management and more. Interface is the result where the client and users can view on the screen. The interfaces will communicate with the business intelligence part in order to output the wanted information.

### F. Data Management

It is important for every organisation to have a sustainable data management plan to ensure that data can be kept up to date and saved. This will help to maintain data quality and availability. It is also vital to archive and backup data in a set time interval to avoid any losses due to data loss or breach of security to protect the business operations. Data management is usually done by the administrators and certain technical employees throughout the data processes up to the interfaces.

### G. Security Management

Data is a crucial component in any organisation. Hence, they need to be protected and maintained at all times. A well-secured organisation would only allow authorized and recognized personnel to access the operational part of the organisation. Hence, firewalls and authentication are highly required in order to protect the data used by Valve. It is also crucial to monitor the activities of the personnel involved in the business operations in order to find out the culprit when there is a breach of security or data loss through logs. This is because Valve is a big company, comprising hundreds of employees and, which suggests that anything could happen.

## V. DATAWAREHOUSE AND OLAP MODEL

The dimensional model serves as the foundation for all analytics and reporting. It is used throughout the entire science of dimensional modelling, not only in the construction of the model but also in the execution of reports and queries, the development of extract, transform, and load tools, and the use of business intelligence tools [13]. Even though the business people in the organisation might never get to see the actual data models that are developed, they would become very acquainted with the reports and dashboards they help produce. Unless those models are capable of generating transparent, effective reporting and analysis, they will be ineffective in assisting the company in viewing data and using it to make informed choices that have an impact on operations. The ability to generate this vital business information is dependent on the use of dimensional modelling.

The dimensional model provides a logical data model for the presentation level of a Data Warehousing and Business Intelligence (DWBI) application, through which the end-user dashboards would take data. It contains a list of the entities and attributes required for the envisioned dashboards. The entities that provide measures are referred to as facts [12][13]. It is termed dimensions when they provide qualifiers that allow facts to be disaggregated, filtered, and organised in various ways. The Steam Sales Schema's data warehouse architecture is centred on the classic star model. Compared to the Snowflake and Starflake schema models, the star schema design has been the most useful when dealing with historical information since it has the lowest query complexity. This allows them to operate at peak performance in data warehouses, data marts, business intelligence applications, and online analytical processing (OLAP).

The fact data can be organised in a fact table at the model's centre, whereas the dimensional data is organised in dimension tables and surrounds the fact table. The fact table is the integration component at the heart of the data warehouse's star structure. They allow machine learning algorithms to analyse the data in its totality and concurrent access to the data by other subsystems. The star schema's dimension tables have no foreign keys. This allows star schema databases to be optimised for lookup and query performance using actual dimensions. They may also be customised to give the best performance and the specific parameters that are deemed most important or often queried by the organization in question. Data may be inputted transactionally as it is received, or it can be imported in bulk and subsequently validated and correctly denormalized.

The fact table of Boeing's dimensional model is the accident fact table, which holds details of the accident like the accident date, publication date, latitude, longitude, number of injuries by type, weather condition when the accident happened, flight phase during the accident and the report's status. From the fact table, it is connected to three dimension tables, namely, the flight dimension, state dimension and the airport dimension.

In the flight dimension, it has an ID number to identify the flight, along with the flight's schedule, purpose and number of passengers. It connects to the aircraft dimension, which records the aircraft's damage, category and identifies whether it is amateurly built. From the aircraft dimension, it connects to two dimensions. The first one is the airline dimension, which identifies the airline owning the aircraft, along with their name, country, region, email and contact number. The second dimension is the model dimension, which identifies the model of the aircraft, with the make, number of engines, engine type and description provided by the Federal Aviation Regulations (FAR). The airline dimension is connected to the manufacturer dimension, which identifies the manufacturer's details and contact methods.

The airport dimension holds information about the flight's departing airport, with details like the airport code, name, country, region, email and contact number. On the other hand, the state dimension recalls where the accident happened, along with the state name. It is connected to the country dimension,

which holds information like the country name and the country's region.

These nine tables make up the dimensional model designed for The Boeing Company. It enables the company to keep track of aviation accidents with paper reports and BI dashboards, while updating it with new records regularly.

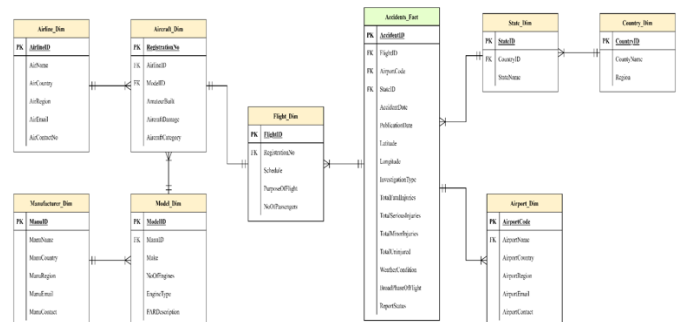


Fig. 24. Diagram of data warehouse schema

The dimensional model that was developed for The Boeing Company uses the snowflake schema as depicted in Fig. 24. Due to the complexity of records to be kept for an aviation accident, the dimensional model needs to use the snowflake schema in order to avoid data redundancy and ensure relevant data is stored in their respective tables. This is possible through the normalization of the dimensional structure, which causes the tables to form this snowflake-like shape, where the name comes from.

## VI. DASHBOARD VISUALIZATION

This project calls for the development of five different dashboards, each of which will be based on an analysis of either aircraft MAKE as shown in Fig. 25, purpose of flight, injury, weather condition or aircraft engine, respectively. Through the use of maps and graphs, data visualization enables us to better comprehend the significance of the information by putting it in a visual context. Due to the fact that the information is simpler to comprehend by the human mind, it is thus far less difficult to recognize trends, patterns, and outliers in massive data sets.

### A. Dashboard for Aircraft MAKE Analysis

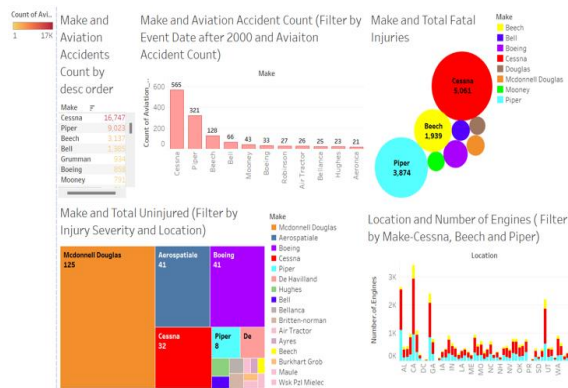


Fig. 25. Dashboard for aircraft MAKE analysis

The term “dashboard” refers to an electronic tracking tool that is used by companies to show and summarize the data that they have collected. The figure above demonstrates the overview of the dashboard showing factors with the “Make” column. This column represents the brand type of the aircraft. As evident from the above dashboard, the Make data is compared with aviation accidents total count, and another similar case where it is also filtered by event date of accidents that occurred after the year 2000. It is also visually represented along with total fatal injuries, denoting only the major ones. Total uninjured is also compared where it is filtered by severity of the accident and location. Final set of comparison has been done taking the top 3 makes having the most fatal injuries and representing them in a bar graph based on location and the number of engines. The dashboard gives an overview of the idea that can be gained by analyzing the representations.

On comparison of make with the count of aviation accidents, it can be found that Cessna had the highest count of accidents being 16,747, which is followed by Piper (9,023) and Beech (3,137) shown in Fig. 26. Evidently, when make is compared with total fatal injuries, Cessna has the highest rate with it being 5,061 and Piper and Beech following the same previous pattern having 3,874 and 1,939 fatal injuries, respectively.



Fig. 26. Make and aviation accidents count by descending order

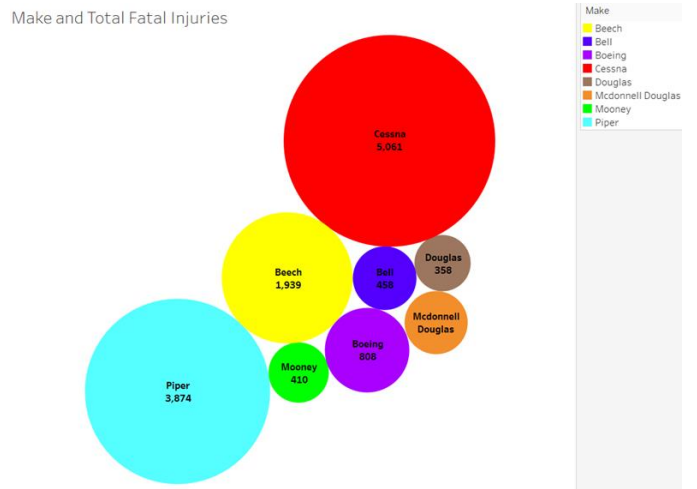


Fig. 27. Make and total fatal injuries

The two visual represents go hand in hand with each other and can aid the company in identifying the makes that have the highest rate of accidents and thus dig deeper into finding the reasons considering other real-life factors into account. It can also be noted that in the first representation, the number is arranged in the decreasing order with the color gradient being darker for the highest number and fading as it lowers. These tools are great for having a more toned representation of data. In the comparison with total fatal injuries, the packed bubble representation gives out a very clear and easy to understand pictorial representation with different makes being color-coded and the higher number having a larger bubble with the count represented as label in Fig. 27.

Based on the insights gained by the previous comparisons, the data can be filtered according to the needs of the user to display certain factors. For example, the tri-colored bar chart has specifically been filtered by the top 3 makes involved in the highest accidents (Cessna, Piper and Beech) and compared with the location of the accident and the number of engines present in the aircraft as shown in Fig. 28.

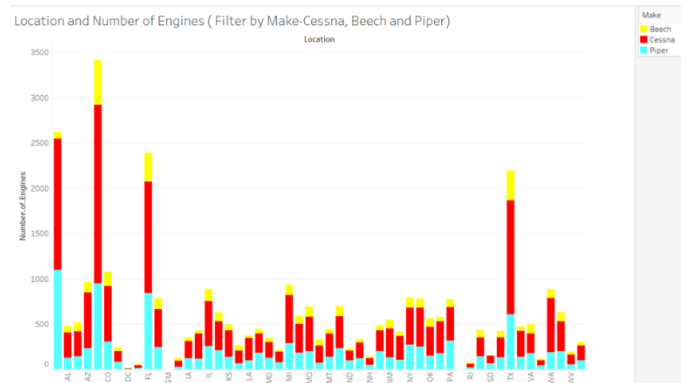


Fig. 28. Location and number of engines, filter by make

Each color represents one of the makes and exhibits a very organized pattern as information. By making use of previously identified information and viewing such representation based on specific requirement, more focused insights can be



obtained. For example, it is proof that Arizona (AZ) has the highest number of accidents happening with the three makes.

Make and Aviation Accident Count (Filter by Event Date after 2000 and Aviation Accident Count)

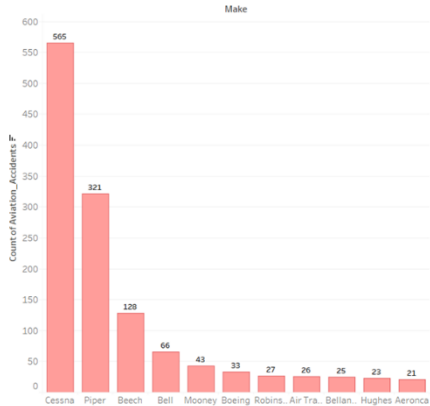


Fig. 29. Make and aviation accident count, filter by event date

Another example can be the bar chart in Fig. 29 depicting the make with the aviation count which is filtered by data to only show those that happened after the year 2000 and by already knowing the major values from the text tables, it has been filtered to show only those with count 21 and above as these alone give a large contrast between the highest values. This allows viewing specific data as per requirement and constraint. In this case, it can be seen that Cessna takes the lead of accidents even in this decade.

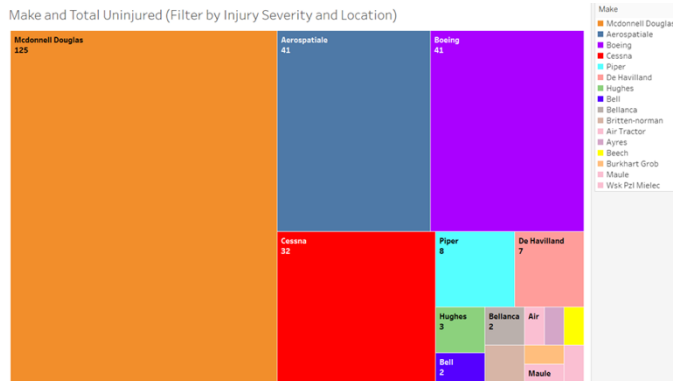


Fig. 30. Make and total uninjured, filter by severity and location

Finally, the Fig. 30 tree-maps present a colorful view of the make and total uninjured in each case that has been filtered by injury severity and location. In this example, two different categories have been allocated as filters with each allocation allowing the flexibility to choose the type and amount of data to be filtered with.

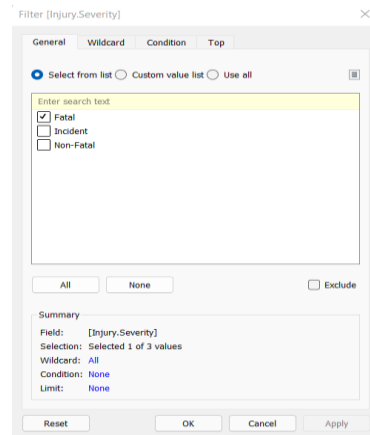


Fig. 31. Filter severity

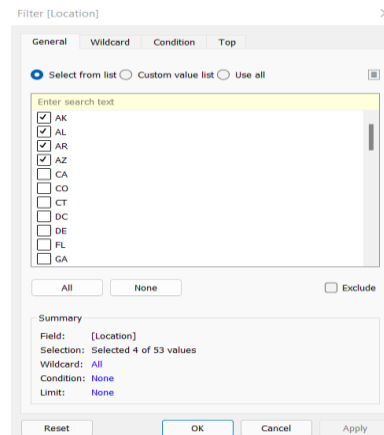


Fig. 32. Filter location

In this case, the severity has been filtered by selecting ‘fatal’ and the location has been restricted to Alaska (AK), Alabama (AL), Arkansas (AR), Arizona (AZ) as shown in Fig. 31 and Fig. 32. Using only these constraints, the output has been presented. This is usually effective when deep analysis has to be made for a specific area or with a unique set of categories.

**B. Dashboard for Purpose of Flight Analysis**

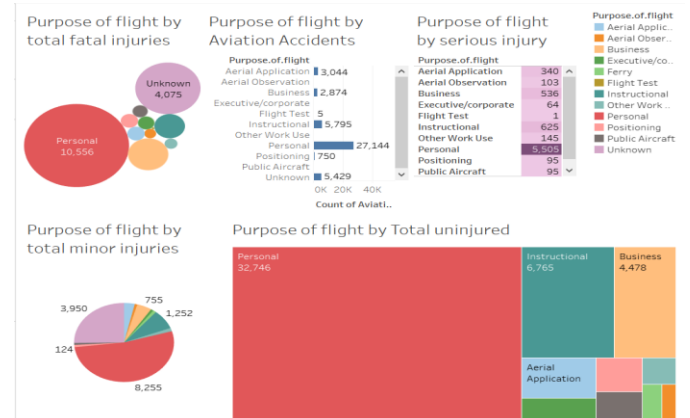


Fig. 33. Dashboard view for purpose of flight analysis

The Fig. 33 represents the dashboard that compares the purpose of flight with multiple factors present in the dataset. The dashboard is useful in providing an overall view of all the insights that have been gathered which can be further used to improve the current situation and identify relations between different factors. In this case, the factor has been compared with total fatal injuries, aviation accidents, total serious injury, total minor injuries and total uninjured.

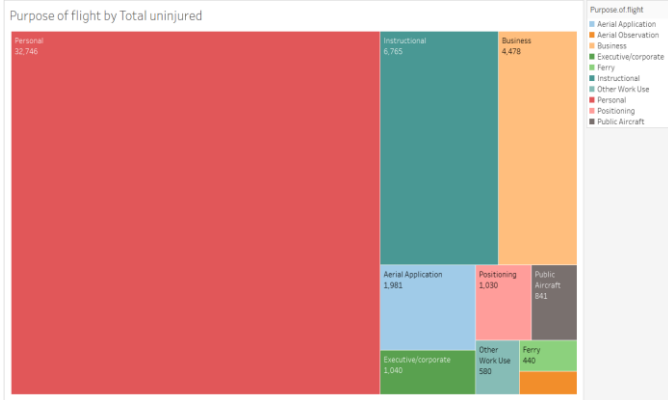


Fig. 34. Tree map for purpose of flight by total uninjured

The Fig. 34 tree-maps represent, the purpose of flight with total uninjured. It can be seen that majority of the flights were due to personal reasons and they also have the highest number of total uninjured people (32,746). With this visual presentation of data, some important information can be collected. In this case, the major reason for travel can be identified, along with the least factor that causes people to travel which is aerial observation.

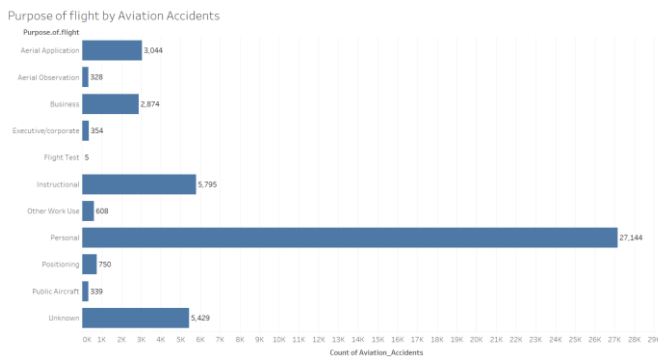


Fig. 35. Bar graph for purpose of flight by aviation accidents

The Fig. 35 bar chart presents the purpose of flight with aviation accidents count. It can be noted that the highest number of accidents is 27,344 and those were people travelling for personal reasons. This is very much related to the previous comparison as both cases denote majority of the people involved in the accidents, travelled for personal use and therefore, they also contribute to a higher number of people uninjured. Bar chart gives a clear and defined view of the data in a very cleansed manner.

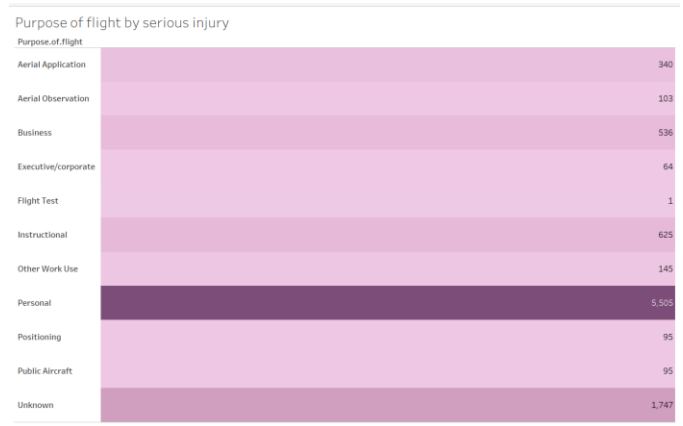


Fig. 36. Bullet graph for purpose of flight by serious injury

The Fig. 36 bullet graph shows the relation between purpose of flight and the number of serious injuries. The highest number of serious injuries is 5,505 which is among people travelling for personal use. As the insight gathered before, we can here specifically identify injury level among passengers travelling with different purposes.

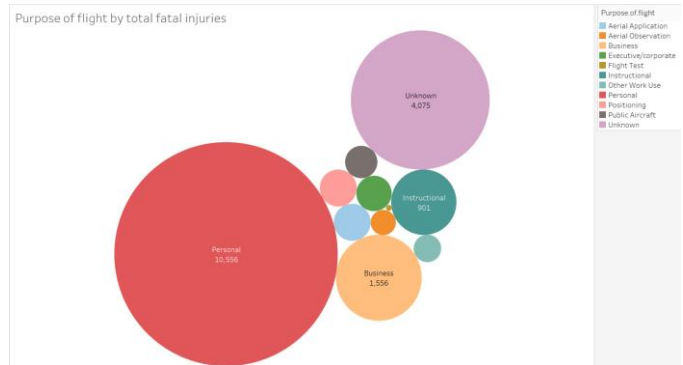


Fig. 37. Packed bubble graph for purpose of flight by total fatal injuries

The Fig. 37 bubble chart shows the amount of people who have been fatally injured due to the accidents and the reasons they travelled. Out of the many categories, many people who travelled for personal and business reasons had fatal injuries. A total of 10,556 injured people were personal travelling passengers while there were around 1,556 business passengers. In this case, the size of the bubble presents the fatal injury count whereas the color represents the different categories of passengers. It shows a very precise and easy to understand view of the data for insight gathering.

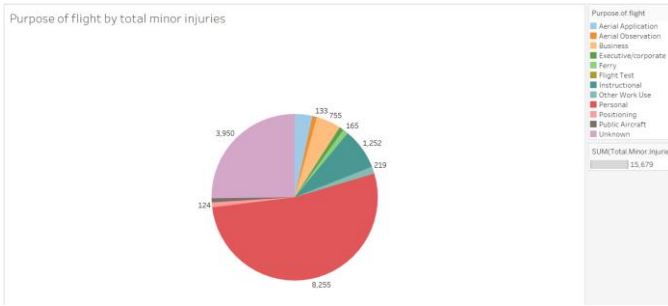


Fig. 38. Pie chart for purpose of flight by total minor injuries

The Fig. 38 pie chart represents the purpose of flight compared with total minor injuries. The different colours depict different categories of passengers while the space occupied by a colour represents the quantity of the people injured. As evident from the previous visual representations, many numbers of people were traveling for personal purposes and therefore, people from that category had the most numbers of injured and uninjured people. In this case, there was a total of 8,255 people from the personal travel category who were had minor injury. The above depictions show the identification of the highest category of people travelling and a deeper insight into the accident and injury data for those different categories.

### C. Dashboard for Injury Analysis

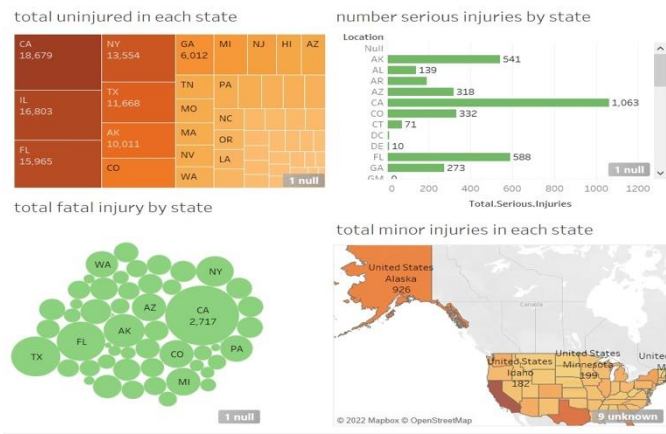


Fig. 39. Dashboard for injury analysis

The Fig. 39 illustrates an overview of the dashboard showing multiple factors from “Location” column. The location column shows the factors according to each state. In this case I used 4 factors named ‘total uninjured, total serious, total fatal and total minor’. Each of which is represented using four different kinds of figures. The dashboard talks all about injuries and deaths. This will help the airline companies to improve their security measures in each state according to the data. Fact Bubbles easily identify where the factor is in high value. Maps are easy to identify geographically. Tree maps can easily give scores for each state and is easy to identify using colour scheme too. Bar charts give accurate statistics of the factors. Following is the explanation of each:

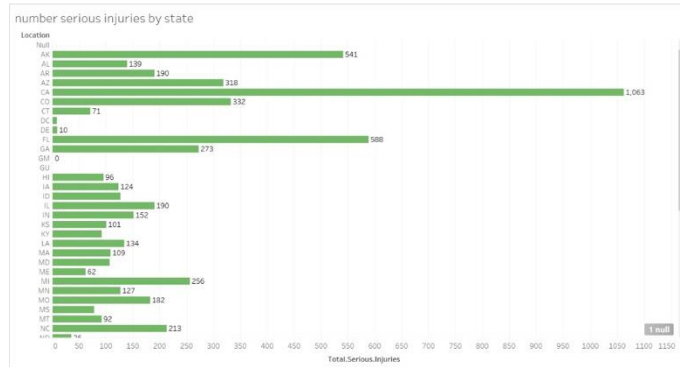


Fig. 40. Bar graph for number of serious injuries by state

According to this Fig. 40, we can analyze that in US, most of the serious injuries are in state CA (California) which are almost close to 1000. After that second highest injuries are in FL and AK. Meanwhile GM and GU have zero serious injuries.

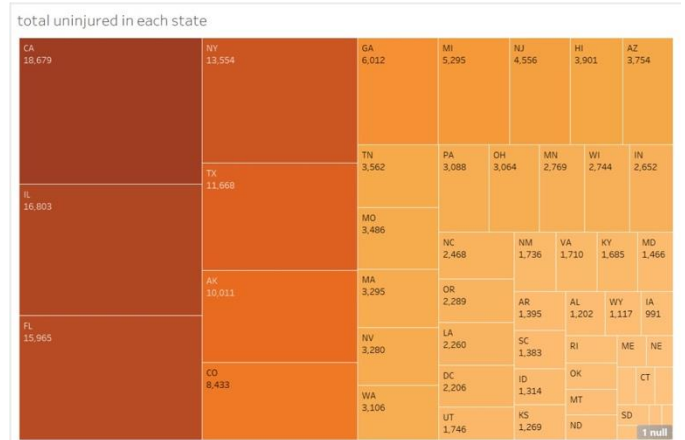


Fig. 41. Tree map for total uninjured in each state

The Fig. 41 relates to uninjured people in each state. We have seen that California had most serious injuries in previous figure but on the other hand it also has most injured people among other states too. IL, FL, and NY are on 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> ranks respectively, each having more than 10,000 people uninjured.

Now coming to the injuries which almost led to death, California is again ranked 1<sup>st</sup> with 2,717 fatal injuries. This might be because of excess number of travellers in CA. TX again is in 2<sup>nd</sup> position with more than 1200 fatal injuries and FL is ranked 3<sup>rd</sup>. Other states have very less reported fatal injuries due to less travellers as shown in Fig. 42.

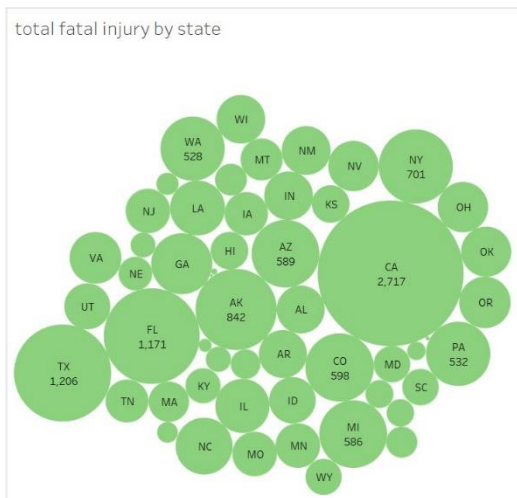


Fig. 42. Packed bubble graph for total fatal injury by state

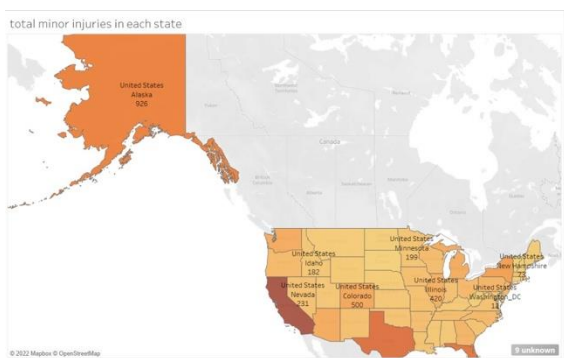


Fig. 43. Map for total minor injuries in each state

Coming to the minor injuries in each state, we used maps to locate them out. The region with most minor injuries is Alaska, with over 900 minor injuries confirmed. Second is Colorado with exactly 500 minor injuries and then ranks Illinois with 420 minor injuries as shown in Fig. 43.

D. Dashboard for Weather Condition Analysis

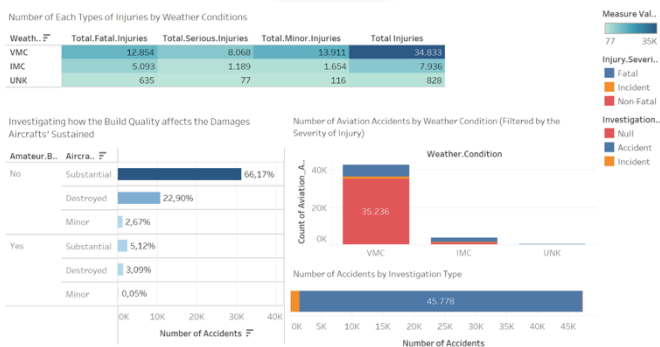


Fig. 44. Dashboard for weather condition analysis

The Fig. 44 displays a dashboard that compares data based on the chosen criteria. In this dashboard, five pieces of information are shown. Three pieces of information used weather conditions as the factor to be compared to another criteria. One other displayed information of damage sustained based on the quality of the builder, which can be seen in the

bottom left. The last one, investigation type is compared with the number of accidents that happened. One of the benefits of a dashboard is that it provides us with the means to view and analyze the key data that has been constructed. Hence, a conclusion can be drawn, and plans can be made accordingly.

Number of Each Types of Injuries by Weather Conditions

Weather	Total.Fatal.Injuries	Total.Serious.Injuries	Total.Minor.Injuries	Total.Injuries
VMC	12,854	8,068	13,911	34,833
IMC	5,093	1,189	1,654	7,936
UNK	635	77	116	828

Fig. 45. Table for number of each types of injuries by weather conditions

The Fig. 45 represents the number of injuries based on the weather conditions. It is displayed in the form of a highlighted text table. The table is constructed to enable the engineers to understand the weather situation that leads to accidents and what types of injuries it leads to. The total injuries caused by a particular weather condition can be analyzed too. Example, the above table shows that VMC generates a higher number of injuries compared to IMC on every level. From there, it can be concluded that humans tend to make more errors even with the situation is favorable. This is because IMC technically describe poor conditions, so machines are more heavily used. Therefore, the engineers can build parts that can improve the performance of humans (pilot and co-pilot).

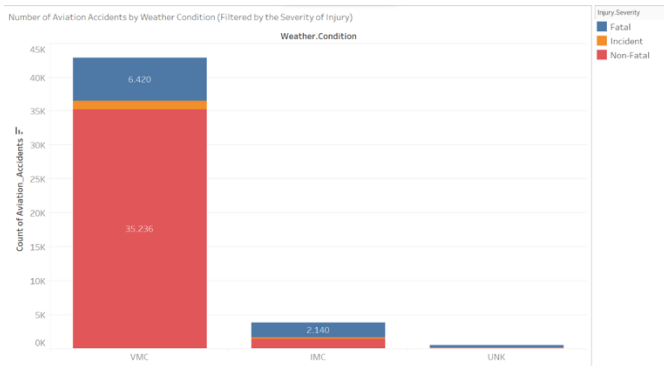


Fig. 46. Stacked bar graph for number of aviation accidents by weather condition

The Fig. 46 bar chart shows the number of accidents caused by weather conditions, where the bar chart is stacked with information on the severity of injury. The severity of injury inside the bar chart is separated with colour marks so the data can be easier to understood. In addition, the number of accidents per severity of injury is also displayed. The data visualization has a similar purpose to the previous one. The difference is that in this data, it describes a more detailed representation of the injury.

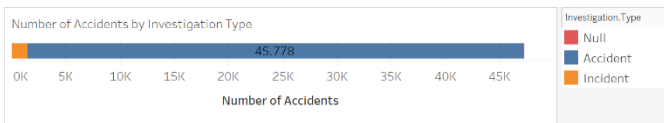


Fig. 47. Stacked bar graph for number of accidents by investigation type

The Fig. 47 stacked bar chart shows the type of accidents. It illustrates the differences in number of each accident type by separating it inside the bar. To differentiate between the

investigation types, colour marks are used. Moreover, the number of accidents for each accident type is also displayed.

There are two reasons why stacked bar is used. The first is to fit the dashboard. The second is because we are dealing with small data where the factor is only one. For efficiency purposes, the number of accidents attribute is put in the column side. This data can be used by engineers to understand the fatalities of the events. An accident is described as an unexpected event that leads to damage and injury. While incident is described as an unexpected event that leads to some minor injury.

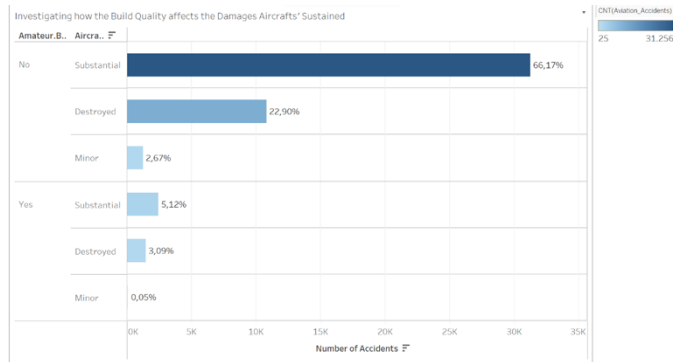


Fig. 48. Graph for investigating how the build quality affects the damages aircrafts' sustained

The data above is visualized using a bar chart in Fig. 48 that displays the damage aircraft sustained because of the build quality. The graph is constructed through two dimensions; the build quality and aircraft damage, and compare it by the number of accidents. The build quality is divided into two parts that compromise the components of aircraft damage. Colour is included to indicate the volume of the data in each criterion. The percentage in the chart above is over the total accidents.

The purpose of the above chart is not to compare directly the number of aircraft damages between aircraft that are maturely built or not. This is because the data distribution between the two criteria is too far. Hence, the above can be used to show the damage distribution on the aircraft. In each build quality, the aircraft mostly sustained substantial damage. It can also be noted that rarely aircraft suffered minor damage during aviation accidents.

### E. Dashboard for Aircraft Engine Analysis

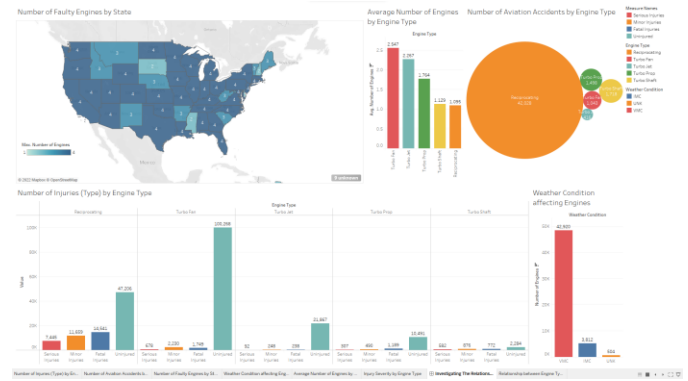


Fig. 49. Dashboard for aircraft engine analysis

Fig. 49 shows a dashboard, it uses information about the aircraft's engines to perform visualizations. In here, there are five different graphs, all designed to help engineers inside of The Boeing Company to get statistics about their aircraft's engines. All these five graphs are there to help engineers and analysts in The Boeing Company to identify faults in their engine's designs or implementation through the performance and safety data. With this information, improvements or changes can be made to the aircrafts produced by Boeing in an effort to increase the safety of the flights and ensure the security of their flight crew and passengers.

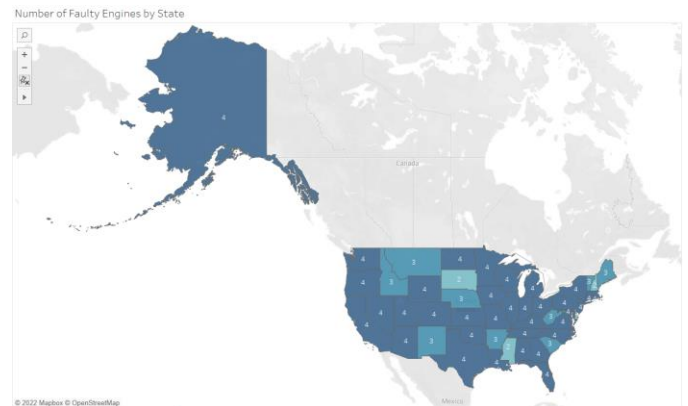


Fig. 50. Map visualization for the number of faulty engines by state

The first graph in the dashboard is a map that Fig. 50 illustrates the number of faulty engines for aviation accidents from each state in the US. It is done by using the state data, along with the maximum number of engines in the accident records for that particular state.

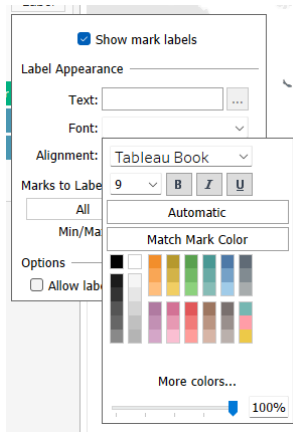


Fig. 51. Settings for label on map

At first, the map only provides an illustration of the number of faulty engines through the shade of the blue colour, being the darker the blue, the more engines that are faulty for that state. However, we thought that this is unclear and would not be enough for the analysts and engineers inside Boeing. Thus, the “Show mark labels” option was enabled as shown Fig. 51, with the label’s text colour adjusted to white, as a black font might be hard to read with dark blues.

This map graph is created so the engineers in Boeing can get to know the maximum number of engines that turned out to be faulty for each of their aircrafts. It is supported with the map to differentiate the data by state, so that in the future, if the departure point of the flight is available for Boeing’s BI dashboard, the engineers can measure the distance between the departing point and point of accident, while investigating the effects from the number of engines used.

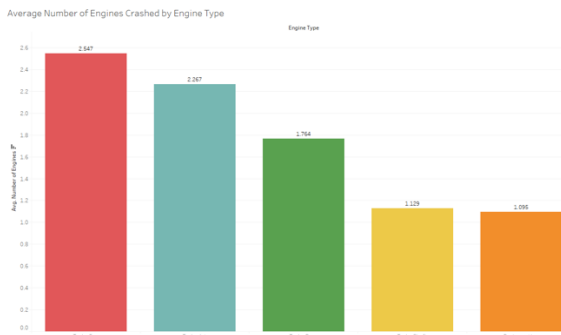


Fig. 52. Bar graph of the average number of engines crashed by engine type

For the second visualization, Fig. 52 it is a bar graph which shows the average number of engines in aircrafts that crashed, separated by its engine type. The different types of engines are shown with the colour mark, which allows for easy reading of the visualization and data. The average number of engines by engine type is also shown, as a label, on top of each bar.

The bar graph would help the aircraft engineers understand the average number of engines installed for each engine type which was involved in aviation accidents. With this data, the engineers can perform a more thorough investigation on the

relationship between the number of engines installed, the installed engine type, its performance and reliability.

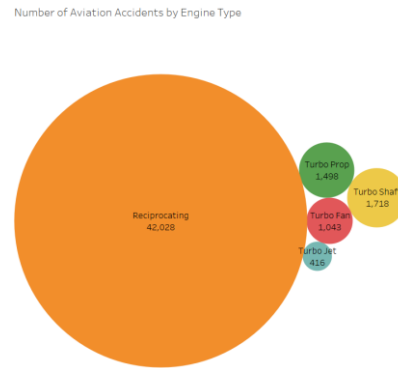


Fig. 53. Packed bubble visualization for the number of aviation accidents by engine type

The third visualization for the engine dashboard is the packed bubble graph as shown in Fig. 53, which shows the number of aviation accidents by its engine type. The packed bubble graph illustrates the difference in accidents between each engine type by differentiating each bubble’s size. The bubbles are also identified with the colour marking option, to make it more easily readable.

The purpose for comparing the number of aviation accidents caused by each engine type through this packed bubble graph is to allow for an easier to understand visualization of the relationship between these two data. The data analysts and engineers can look at this graph and get to know the number of accidents involved by each different engine type. For example, the packed bubble graph shows that the reciprocating engine type caused the greatest number of aviation accidents. Thus, they can investigate into the reason behind the engine type’s performance, to determine whether improvements should be made, or a different engine should be used.

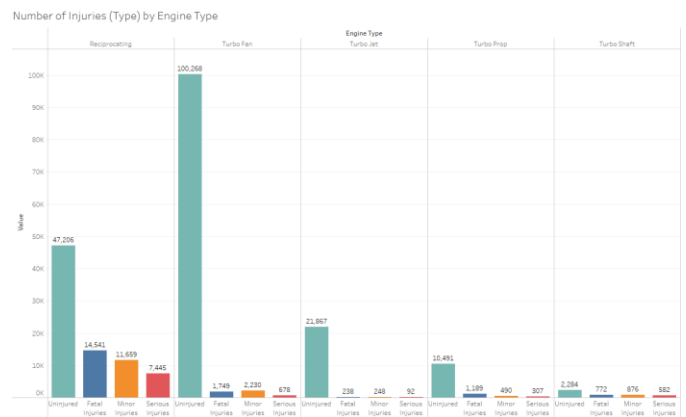


Fig. 54. Bar graph of the number of injuries (type) by engine type

The fourth visualization is shown in Fig. 54 bar graph, which shows the difference in number of each injury type, separated by the engine type. This graph is possible by putting the total number of fatal injuries, serious injuries, minor injuries and uninjured into one measure values group, then comparing it with the engine type. Each engine type is divided

into one section, with five sections in total. While the number of injuries by type are shown for each engine type using different colours for each injury type.

This bar graph can be considered as the one that ties all of the previous visualization together. This is because it shows the total for each type of injury, compared among different engine types. Previously, we had a bar graph which showed the average number of engines for each engine type and a packed bubble graph to illustrate the number of aviation accidents by engine type. From these two graphs, the engineer or the analyst inside Boeing may conclude that the low number of average engines is one of the reasons that reciprocating engine types were involved in most aviation accidents. While the turbo fan engine type used, on average, the most engines per aircraft, at the same time having the second least aviation accidents.

This bar graph goes further to help the analysis by showing that, while luckily most of the passengers involved were uninjured, the reciprocating engine type had the greatest number of injuries for every injury type. On the other hand, the high average number of engines used for turbo fan aircrafts attained the greatest number of uninjured passengers. Judging from the data on the dashboard, the engineers may try to incorporate characteristics or improvements of turbo fan engines into reciprocating engines, in hopes to improve the safety and reliability of aircrafts with the engine type. Otherwise, the engineers can phase out reciprocating engine aircrafts in favour for safer engine types.

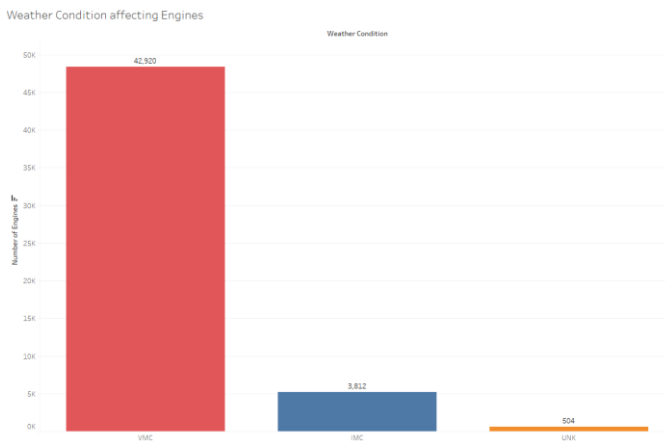


Fig. 55. Bar graph for the weather condition affecting engines

The last visualization for this engine dashboard is shown in Fig. 55 the bar graph which shows the how the different weather conditions affect the aircraft's engines. It is done by color coding the different weather conditions and comparing them with the accident numbers with the engine types.

In aviation, VMC means that the weather condition is good enough that the pilot can attain visuals of the skies, nearby possible aircrafts and the nearby terrain. Meanwhile, IMC describes when the visibility of the aircraft's surroundings is limited by bad weather conditions or when flying through thick clouds. In a situation like this, trained pilots will have to rely on the hearing equipment available in the cockpit to

remain updated about the aircraft's current whereabouts and surroundings.

This current state of the graph eliminates one common suspicion of engineers, which is that bad weather conditions are the main reasons that aviation accidents happen. However according to this bar graph, it is not the case as most of the accidents recorded actually happened when the weather condition is satisfactory for flight. While bad weather conditions do also play considerably small part in affecting the chance of an aviation accident happening, pilots nowadays have access to more modern and advanced monitoring equipment that allows them to keep track of the plane's status, location, elevation and such. This lends into the argument that weather conditions do not necessarily cause aviation accidents, but instead fault in the aircraft's design or the pilot's inexperience or mistake had caused the accident to happen.

## VII. CONCLUSION

BI is put into action using programming or devices which assemble tremendous measures of information and change them into valuable data which can be investigated to likewise acquire bits of knowledge and decide. Utilizing this methodology, the association will actually want to recognize past patterns and information that untruths concealed inside verifiable information gathered throughout the long term.

Boeing being the main organization in the flying area has numerous obligations and should stand against all the challenges it is facing right now. Most of the challenges include unexpected government regulations, market demands; technological advancements market demands, global coordination risk and preapproval tests. Boeing can combat all these problems easily by working on the market opportunities. For instance, it can increase its seating capacity and produce engines with more fuel efficiency as compared to competitors. Moreover, the Boeing team can analyze past incidents and an effective design change considering past accidents.

Finally, it can improve its technological side by doing some advancement in exchange of digital information, weather prediction, connected cabin systems and cameras to provide passengers a view of outside plane. Coming to the solutions, Boeing can implement prediction of encountering into a crash using BI and then use countermeasures to prevent it from happening. Another application where BI can be used is aircraft maintenance scheduling using predictive maintenance. Business Intelligence is also extremely beneficial in identifying the cause of aircraft crash. The autopilot system also learns from such situations and prevents the craft from being going into deadly situations too.

To sum this up BI is an essential tool in modern day businesses and organizations. It helps companies falling from mishaps and increases their performance along with user security. The data was clearly assessed in this assignment, and we got valuable statistics from it which can also help the company improve its insights.

## REFERENCES

- [1] Boeing (2019). *Boeing: Data Analytics*. [online] [www.boeing.com](http://www.boeing.com). Available at: <https://www.boeing.com/services/government/data-analytics.page>.
- [2] Calzon, B. (2022). *Discover Data Warehouse & Business Intelligence Architecture*. [online] BI Blog | Data Visualization & Analytics Blog | datapine. Available at: <https://www.datapine.com/blog/data-warehousing-and-business-intelligence-architecture/>.
- [3] HT (2019). *Naive Bayes Algorithm*. [online] Medium. Available at: <https://medium.com/@hackares/naive-bayes-algorithm-e565daa89eb7>.
- [4] IBM Cloud Education (2020). *What is Data Modeling?* [online] [www.ibm.com](http://www.ibm.com). Available at: <https://www.ibm.com/cloud/learn/data-modeling>.
- [5] Kaggle (2022). *Aviation Accident Database & Synopses*. [online] [kaggle.com](http://kaggle.com). Available at: <https://www.kaggle.com/khsamaha/aviation-accident-database-synopses>.
- [6] KDnuggets (2020). *Decision Tree Algorithm, Explained*. [online] KDnuggets. Available at: <https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html>.
- [7] Mark Logic (2022). *Boeing: Querying Model Based Systems Definition Data | MarkLogic World*. [online] MarkLogic. Available at: <https://www.marklogic.com/resources/boeing-querying-model-based-systems-definition-data/>
- [8] Mihai, A. (2015). Airline Applications of Business Intelligence Systems. *INCAS BULLETIN*, 7(3), pp.153–160. doi:10.13111/2066-8201.2015.7.3.14.
- [9] Rushikesh Pupale (2018). *Medium*. [online] Medium. Available at: <https://towardsdatascience.com/https-medium-com-pupalrushikesh-svm-f4b42800e989>.
- [10] Saud, D. (2021). *How can business intelligence revolutionize the airline industry*. [online] Folio3 Dynamics Blog. Available at: <https://dynamics.folio3.com/blog/business-intelligence-revolutionize-airline-industry/>.
- [11] Segal, E. (2022). *Boeing Faces New Challenges To Image, Reputation And Credibility*. [online] Forbes. Available at: <https://www.forbes.com/sites/edwardsegal/2021/10/15/boeing-faces-new-challenges-to-image-reputation-and-credibility/?sh=2d1b3fbb6bff>
- [12] Sherman, R. (2020). *7 Data Modeling Techniques and Concepts for Business*. [online] SearchDataManagement. Available at: <https://www.techtarget.com/searchdatamanagement/tip/7-data-modeling-techniques-and-concepts-for-business>.
- [13] Taylor, D. (2022). *What is Dimensional Modeling in Data Warehouse? Learn Types*. [online] [www.guru99.com](http://www.guru99.com). Available at: <https://www.guru99.com/dimensional-model-data-warehouse.html>.
- [14] Tripathi, A., Bagga, T. and Aggarwal, R.K. (2020). Strategic Impact of Business Intelligence : A Review of Literature. *Prabandhan: Indian Journal of Management*, 13(3), p.35. doi:10.17010/pijom/2020/v13i3/151175.



# Metaphor Recognition Method based on Graph Neural Network

Zhou Chuwei<sup>1</sup>, SHI Yunmei<sup>2</sup>

Beijing Information Science and Technology University, Beijing Key Laboratory of Internet Culture Digital Dissemination,  
Beijing 100101, Beijing, China<sup>1,2</sup>  
School of Computer, Beijing University of Information Technology, Beijing 100101, Beijing, China<sup>1,2</sup>

**Abstract**—This Metaphor is a very common language phenomenon. Human language often uses metaphor to express emotion, and metaphor recognition is also an important research content in the field of NLP. Official documents are a serious style and do not usually use rhetorical sentences. This paper aims to identify rhetorical metaphorical sentences in official documents. The use of metaphors in metaphorical sentences depends on the context. Based on this linguistic feature, this paper proposes a BertGAT model, which uses Bert to extract semantic features of sentences and transform the dependency relationship between Chinese text and sentences into connected graphs. Finally, the graph attention neural network is used to learn semantic features and syntactic structure information to complete sentence metaphor recognition. The proposed model is tested on the constructed domain dataset and the sentiment public dataset respectively. Experimental results show that the method proposed in this paper can effectively improve the recognition ability of metaphorical emotional sentences.

**Keywords**—Sentiment analysis; metaphor recognition; graph neural network; attention mechanism

## I. INTRODUCTION

The use of metaphorical sentences can make the described things more imaginative and vivid, and can show stronger emotions. However, official documents are issued by legal organs or organizations, and their contents should be clear and specific, and their expressions should be simple and easy to understand, so they do not need to deliberately pursue rhetoric and generally do not use metaphors, lyricism and symbols. Metaphor is a common rhetorical technique in metaphorical sentences, but it is generally not used in official documents for standardization. It is a time-consuming and laborious task to manually check metaphorical sentences in official documents, and using automatic detection tools can greatly reduce the workload of document writers and increase the accuracy and efficiency of identification.

Metaphor sentence recognition is the identification of the presence or absence of metaphorical usage in a sentence, which is essentially a dichotomous classification task. There are rule-based methods, statistical machine learning methods and deep learning methods for metaphor recognition. Using rule-based approaches [1][2] can effectively address simple, common metaphor types, but requires a lot of manual effort to design rules and does not easily cover the full range of identified types. Using traditional machine learning methods [3][4] can identify more comprehensive types of metaphorical sentences with higher recognition rates, but the method is

difficult to learn deep semantic features and the model generalization is more difficult. In recent years, deep learning methods have become popular, and deep learning-based methods are widely used in metaphor recognition [5][6][7][8][9][10], which are similar to text classification tasks, where text is first transformed into vector form, and then neural network models are used to learn the semantic information of sentences. A better recognition result is usually obtained given enough corpus. In the last two years, the introduction of pre-trained language models [12] has further improved metaphor recognition, but the existing studies do not use syntactic structure information or do not deeply integrate syntactic structure information into the models.

Metaphorical sentences appear in a specific contextual environment, and without the contextual relationship, there is no meaning of metaphorical sentences. Therefore, identifying metaphorical sentences requires establishing a good contextual relationship between words, so as to identify whether a sentence has words containing metaphorical relationships. In this paper, we use Bert to extract the semantic features of sentences, convert the text into word embedding information, then perform syntactic dependency analysis, and integrate semantic information with syntactic structure information when constructing sentence connectivity graphs, and finally use Graph Attention Networks (GAT) to learn the semantic and syntactic structure information of sentences, and fully extract the sentence feature information to identify metaphorical sentences.

The contributions of this paper are as follows:

The proposed BertGAT model for recognizing metaphorical sentences in Kumon is proposed, which incorporates syntactic structure information of sentences based on the external knowledge of the pre-trained model to make sentence modeling more refined.

The proposed BertGAT model can learn the sentence structure information and weight the word nodes with varying importance, which in turn enriches the text feature information.

A rhetorical implicit sentiment recognition dataset for the metaphor class in the public domain is constructed, and the proposed BertGAT model in this paper achieves accurate results not only in this dataset, but also in the public dataset ChnSentiCorp with good results.

In this paper, a metaphor recognition model based on graph neural network and pre-training model bert is proposed.

Therefore, the graph neural network and pre-training model are introduced respectively. Finally, the validity of the model in domain data set and public data set is verified by experiments.

## II. RELATED STUDIES

### A. Metaphorical Sentence Recognition

Recognition of metaphorical sentences is a common and important problem in Natural Language Processing (NLP) and has been studied by researchers for a long time. The earliest metaphor recognition is the rule-based template approach proposed by Krishnakumaran [1], which relies on WordNet, an a priori knowledge base, to identify "A is B" noun metaphors, "verb+noun" verb metaphors, and "adjective+noun" metaphors by using its contextual relations and setting frequency thresholds "noun" verb metaphor and "adj+noun" adjective metaphor. The drawback of this method is that it relies too much on the a priori knowledge base WordNet and the identified metaphor sentences are simple. Noun metaphors are the most common expressions in lines, followed by verb metaphors, so researchers generally focus on the identification of these two types of metaphorical sentences [2][3][5][6][7][8][9][10].

Su Chang [2] proposed four algorithms to identify noun metaphors from different perspectives and recognized rich types of noun metaphor sentences with high accuracy, but the method relied entirely on WordNet. Shutova [6] proposed a clustering-based approach to identify verb metaphor sentences by taking verb subject groups and verb object groups that have been labeled as metaphorical expressions as seed sets and continuously expanding the seed sets by clustering. When these phrases appear in a sentence, it is judged as a metaphorical sentence, and the performance of this method depends on the quality and diversity of the seed words. Su Chang et al [3] mined multi-level features based on multi-level semantic networks, including full-text features, attribute features, perceptual features, etc., but for some complex metaphorical sentences, some attributes of the target domain are difficult to extract.

In recent years, deep learning has been developed and applied more and more widely in the field of NLP. The classical CNN model is not only simple but also powerful, Kim Y et al [4] applied CNN to text classification and achieved good results, while metaphor recognition is essentially text classification, some researchers use CNN models [7][9] in metaphor recognition to capture locally relevant features of text. Different ranges of semantic features can be captured by convolutional kernels of different sizes. Kui-Lin Su [9] used a pre-trained model Bert with external semantic knowledge and fused it with a CNN model for noun metaphor recognition.

BiLSTM (Bi-directional Long Short-Term Memory) is formed by combining forward LSTM with backward LSTM, which has stronger bi-directional semantic dependency capture capability, and also effectively prevents the gradient explosion and gradient disappearance problems of long text sequences by using gate mechanism and memory units. Jiaying Zhu [7] combined CNN and BiLSTM for metaphor recognition, in

which CNNs with different windows were used to obtain semantic features in different ranges, and BiLSTM was used to concatenate local semantic features extracted by CNNs and obtain full-text features from them, and the model could recognize verb metaphors, noun metaphors, preposition metaphors, etc. Chuandong Su et al [8] based on Bert used five features for metaphor recognition to identify whether the sentence and the query words are metaphorically related.

Shenglong Zhang [10] used pre-trained language models to extract sentence semantic features, used Graph Convolutional Network (GCN) to extract syntactic features from syntactic structure information, and finally obtained classification results of sentence metaphors by splicing sentence semantic features and syntactic features classification, which achieved better recognition results on verb metaphors and noun metaphors.

Existing deep learning-based studies make full use of sentence sequence information or incorporate external semantic knowledge from pre-trained models, but do not make full use of the supporting role of sentence structure, and some studies have shown that word dependencies are subject-verb structures or verb-direct object structures, which are more likely to be expressed metaphorically [11]. This shows that syntactic structure is also an important feature for identifying metaphors.

In this paper, we propose a pre-trained language model as a word embedding layer for the recognition of metaphorical types of rhetoric in the public domain, and construct a connected graph using dependency syntactic analysis with nodes embedded in the corresponding word vectors, which enables the graph attention neural network to learn sentence semantic and syntactic structure information more fully and improve the effectiveness of metaphor recognition.

### B. Bert Model

Bert [12] is a pre-trained model proposed by Google AI Research Institute in October 2018. Bert stacks multiple bi-directional Transformer structures and uses residual connections to address the limitations of traditional models in one direction and the problem of long-term dependencies. The use of the pre-trained language model Bert is divided into two phases: model pre-training and model fine-tuning.

In the model pre-training phase, there are two training tasks, namely word masking task and next sentence prediction task, through which the Bert model learns the relationships between words and lays the foundation for downstream tasks.

Bert's excellent performance and strong generalization performance make it widely used in the field of NLP, and many researchers have developed on the Bert model and then proposed improved models based on Bert, which has become one of the most commonly used techniques in the field of NLP.

The Bert model can replace word vectors, and in addition to transforming text into a computer-processable vector, it can further extract the deeper information embedded in the text. The Bert model introduces the relative position information

between words, so it can effectively extract the correlation and difference between ontology and metaphor in metaphor [13].

### C. Graphical Neural Networks

Graph neural network is an algorithm that fuses deep learning and graph structure, and its main idea is to use graph propagation mechanism and then update the features of nodes using deep learning. With stronger interpretability and good at processing graph structured data, graph neural networks are widely used in social systems, transportation networks, knowledge graphs and other fields.

In recent years, graph neural networks have received increasing attention, and the field of NLP has likewise introduced the study of graph neural networks. When graph neural networks are used for text processing, serialized text needs to be modeled first so as to introduce text structure information and then combined with downstream tasks to process text data using deep learning-based graph propagation algorithms. Graph neural networks use words as nodes and can introduce additional information to enrich features, such as dependency relationships [14], co-occurrence information between words [15], etc.

In the field of NLP, there are two common graph neural networks, GCN and GAT, in which the neighboring nodes of a node have the same weight, but the association between nodes

usually has different importance, and GAT improves this by using the attention mechanism. The use of multi-headed attention also prevents overfitting the multi-headed attention mechanism and enhances the expressive power of the network [16].

The graph neural network model for metaphor recognition proposed in this paper introduces the dependent syntactic structure information between words in order to fully integrate the sentence structure information, and makes use of the powerful language structure modeling capability and structured feature extraction capability of GAT in order to learn the dependent syntactic information from the sentence dependent syntactic tree [17].

### III. METAPHOR RECOGNITION MODEL BERTGAT

The metaphor recognition model BertGAT proposed in this paper is divided into 4 layers, which are input layer, coding layer, graph attention layer and output layer, as shown in Fig. 1. First, the text is input to Bert and converted into word embedding information; then the sentence connectivity graph is constructed using LTP, feature extraction is performed using GAT, and finally the features are fed into the classifier to get the probability of each category. The coding layer and the graph attention layer are the core parts of the model.

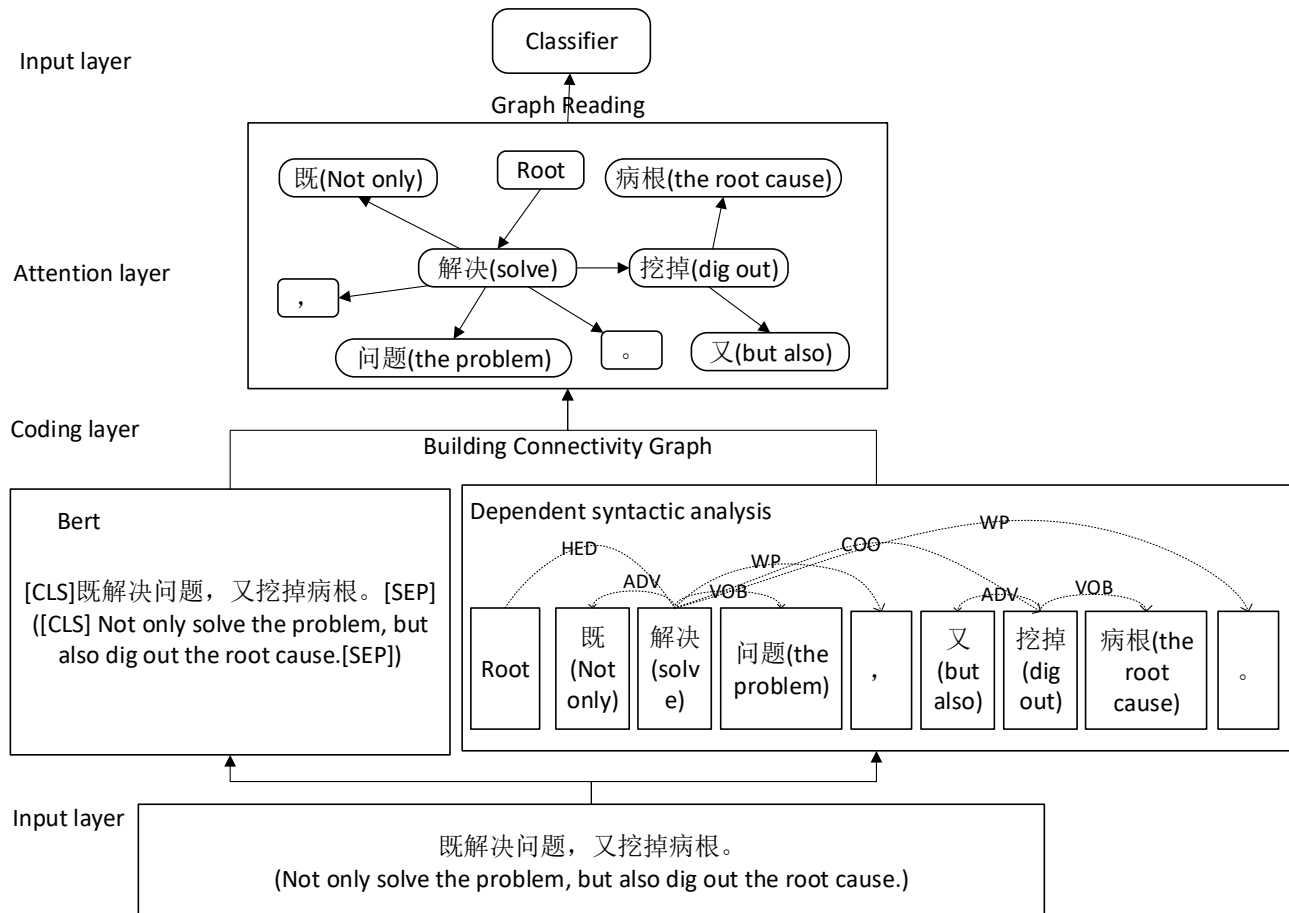


Fig. 1. Structure of BertGAT model.

The coding layer consists of Bert and the dependent syntactic analysis module. In the Bert pre-trained language model, sentences S in the text are sliced into word or character sequences, and after Bert processing, the output is a sequence of word vectors. It has been demonstrated that the dependency syntactic tree helps to improve the recognition rate of sentiment analysis [18][19] because it provides dependency connections between words, which leads the model to better learn the dependency information between long-distance words. In order for the model to fully learn the dependency information between words, the open source LTP tool of Harbin Institute of Technology is used for dependency syntactic analysis to obtain sentence dependencies and their clauses, and the word vectors belonging to the same clause are summed to construct a sentence connectivity graph by combining the dependencies, and the dependencies on each edge are embedded in low-dimensional vectors, and the graph nodes are embedded in word or character vectors.

The graph attention layer processes the connected graph of sentences to fully obtain the syntactic structure and semantic information of sentences, and finally outputs graph representation vectors as sentence features and input them into the classifier.

#### A. Input Layer

The role of the input layer is to perform data pre-processing in preparation for input to the encoding layer.

Before input to the Bert model, the sentence is cut into a sequence of words or characters, as shown in Equation (1).

$$S1 = \{[CLS], c_1, c_2, c_3, \dots, c_n, [SEP]\} \quad (1)$$

[CLS] and [SEP] are special identifiers for Bert, where [CLS] is added at the beginning of the sentence and [SEP] is added at the end of the sentence, and  $c_i$  is the character or word after the sentence is cut.

Chinese word separation is performed before input to the syntactic analysis module.

$$S2 = \{w_1, w_2, w_3, \dots, w_n\} \quad (2)$$

$w_i$  is the word after sentence segmentation.

#### B. Coding Layer

The encoding layer consists of the Bert module and the dependent syntactic analysis module. The pre-trained language model used in the Bert module is the original bert-base-chinese released by Google, containing 12 Transformer coding blocks, which are fine-tuned and trained together with the parameters of the attention layer to achieve metaphorical sentence recognition in Kumon.

Fig. 2 shows the model structure of Bert, which transforms the text sequence [CLS],  $c_1, c_2, c_3, \dots, c_n$  [SEP] into a word or character vector  $e_{[CLS]}, e_1, e_2, \dots, e_n, e_{[SEP]}$ , and the final output is  $x_{[CLS]}, x_1, x_2, x_3, \dots, x_n, x_{[SEP]}$ , which is the vector generated by the encoder capturing the semantics of the multilayer Transformer structure.

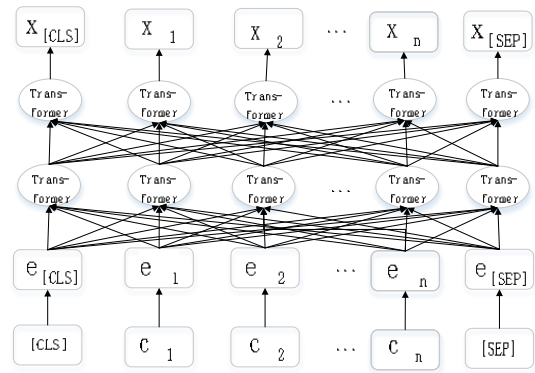


Fig. 2. Bert model structure.

This sequence S1 is input to the Bert model, and the Bert layer converts the text sequence S1 into the form of a vector and captures the semantics of each word or character using the Transform encoder to generate an embedded sequence X containing contextual semantic information, as shown in equation (3).

$$X = \{x_{[CLS]}, x_1, x_2, x_3, \dots, x_n, x_{[SEP]}\} \quad (3)$$

The output after the Bert module is a character-level vector embedding, while the generation of word-level vector embeddings is combined with LTP dependent syntactic analysis.

The LTP tool transforms the sentence sequence S2 into a triplet sequence  $T = \{t_1, t_2, t_3, \dots, t_n\}$ . The triad  $t_j$  is shown as follows.

$$t_j = (g, q, d) \quad (4)$$

where  $t_j$  is the  $j$ th set of dependency information,  $g$  is the dominant word,  $q$  is the modifier, and  $d$  is the dependency.

Finally, Bert is used to generate word embeddings, which are combined with LTP tool dependent syntactic analysis to generate a connected graph representation of the sentences.

The character-level vectors generated by the Bert pre-training model are represented as word vectors by summing the word vectors belonging to the same word.

#### C. Attention Layer

The role of this layer is to mine the sentiment association between word nodes and construct a connectivity graph based on a sequence of triples, where the nodes are word vectors or character vectors and the edges are dependencies represented by onehot. Based on the constructed sentence connectivity graph, the features between nodes are captured using graph attention network learning.

The following shows the node  $i$  at layer  $l$ . The similarity coefficients between it and its neighbor nodes are calculated one by one.

$$a_{ij}^l = \frac{\exp(f(a_i^T [W^l h_i^l] \| [W^l h_j^l] \| d_{ij})))}{\sum_{j \in \{i\}} \exp(f(a_i^T [W^l h_i^l] \| [W^l h_j^l] \| d_{ij})))} \quad (5)$$

The above  $a_{ij}^l$  denotes the layer  $l$ , the weights of node  $i$  to neighbor node  $j$ .  $W^l$  is the weight matrix of the model trained at layer  $l$ .  $h_i^l$  is the word vector representation of node  $i$  at layer  $l$ .  $n[i]$  is the set of neighbor nodes of node  $i$ .  $d_{ij}$  is the dependency between the  $i$  node and the  $j$  node.

The weights between the word nodes and the  $n$  neighbors can be calculated by equation (5).

$$a_i^l = \{a_{i1}^l, a_{i2}^l, \dots, a_{in}^l\} \quad (6)$$

$a_i^l$  identifies the importance level between word node  $i$  and its neighbor nodes, and combined with the node vector of the previous layer, the word node word vector of layer  $l+1$  can be updated.

$$h_i^{l+1} = \sigma(\sum_{j \in n[i]} a_{ij}^l w^l h_j^l) \quad (7)$$

$h_i^{l+1}$  is the word vector representation of point  $i$  at layer  $l+1$ , and  $\sigma$  is the ELU activation function. In a sentence, there are usually key words that determine the sentiment of the sentence, so the attention mechanism can be used to identify these key words and give them a greater weight. In the sentence "to solve the problem but also to dig out the root of the disease", the word "root of the disease" has a more obvious emotional tendency, so the weight value will be higher and the influence on neighboring nodes will be greater.

A study [11] pointed out that syntactic structure helps in metaphor recognition, so the graph constructed based on dependent syntactic analysis with dependencies can add syntactic structure information to enable the model to mine possible metaphorical expressions, in addition to directing the model to focus on important words.

The use of the multi-headed attention mechanism enables the model to mine information from multiple dimensions and prevents the model from overfitting, and the output after the introduction of the multi-headed attention mechanism is shown in the following equation.

$$h_i^{l+1} = \sigma(\frac{1}{K} \sum_{k=1}^K \sum_{j \in n[i]} a_{kij}^l w_k^l h_j^l) \quad (8)$$

$h_i^{l+1}$  is the feature vector of node  $i$  in this layer after feature extraction by the graph neural network with multi-head attention mechanism.  $K$  indicates that  $K$  attention heads are used;  $a_{kij}^l$  is the similarity coefficient between node  $i$  and node  $j$  after calculation by the attention mechanism of the  $k$

head;  $w_k^l$  is the weight matrix of the linear transformation of the input vector at the  $k$  head.

#### D. Output Layer

After the above feature extraction of graph attention neural network, the graph representation vector is obtained by reading out all graph node features using Sum. After the calculation of the multi-headed attention mechanism, the obtained graph representation vector is  $R(H)$  as in the following equation.

$$H = \{h_1, h_2, h_3, \dots, h_n\} \quad (9)$$

$$R(H) = \sum_{i=1}^n h_i \quad (10)$$

$H$  is the whole graph and  $h_i$  is the  $i$  node feature vector of the graph.  $R(H)$  is the readout of the whole graph  $H$  using node Sum.

After obtaining the graph representation vector representing the sentence feature information, a fully connected softmax function is used to classify the sentiment, where the category with the highest probability is the sentiment category of the predicted text, as shown in the following formula:

$$p = \text{softmax}(wR(H) + b) \quad (11)$$

$w$  is the weight matrix of the model training,  $b$  is the bias matrix,  $p$  is the prediction category of the model.

## IV. EXPERIMENT

### A. Authors and Affiliations Data sets and Assessment Indicators

This experiment is oriented to the recognition of sentences containing implicit sentiment in government official documents, for which there is no data available, so we need to construct a domain dataset. We search for metaphorical sentences commonly used in official documents in the Internet, among which are mainly noun metaphor sentences, and combine the verb metaphor and noun metaphor sentences in the Chinese metaphor recognition dataset released in the CCL 2018 Chinese Metaphor Recognition and Sentiment Analysis Task, a total of about 5000 sentences as positive examples; we use a crawler to crawl 30,000 official documents on government official document websites, and randomly select about 5000 sentences as counterexamples. Several sentences in the dataset are listed in Table I.

TABLE I. EXAMPLE TABLE OF METAPHORICAL SENTENCES

Text	Category	Source
切忌眉毛胡子一把抓，防止芝麻西瓜一起抱。 (Do not browbeat, to prevent the sesame seeds and watermelons together to hold.)	Noun Metaphor	Internet
经济转型中，政府必须起到催化剂的作用。 (The government must act as a catalyst in economic transformation.)	Noun Metaphor	CCL 2018
栽培社会所需要的集体主义精神。 (Cultivate the spirit of collectivism that society needs.)	Verb Metaphor	CCL 2018
中国青年是有深厚家国情怀的青年。 (Chinese youth are young people with a deep sense of family and country.)	Literary	Internet

In the first sentence of Table I, eyebrows and beards, sesame seeds and watermelons are metaphors for priorities and sub-priorities in work, which are noun metaphors. These sentences are often used in speeches of leaders to promote their policies, ideas and work, but not in strictly official documents. The second sentence is a noun metaphor, comparing the government to a catalyst. The third sentence is a verb metaphor, where the verb "cultivate" refers to planting and nurturing, but the sentence refers to cultivating.

To further validate the model for sentiment recognition, the public dataset ChnSentiCorp [20] was used to further validate the model effect.

The training set, validation set and test set were assigned in the ratio of 8:1:1.

The result evaluation metrics used recall, precision, accuracy and F1 value formulas as shown in (12) to (15) below.

$$\text{recall} = \frac{TP}{TP+FN} \times 100\% \quad (12)$$

$$\text{precision} = \frac{TP}{TP+FP} \times 100\% \quad (13)$$

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{(\text{precision} + \text{recall})} \times 100\% \quad (14)$$

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (15)$$

### B. Parameter Setting and Experimental Environment

The experimental parameters of the BertGAT model proposed in this paper are set as follows:

1) *Experimental environment*: Using deep learning framework pytorch, operating system ubuntuServer 18.04.5, GPU is Tesla P100, memory 60G, programming language python3.8.

2) *Parameter setting in Bert*: Bert uses the original Bert-base-Chinese released by Google, the dropout is set to 0.3, the optimizer uses AdamW, the maximum number of individual text words is 300, the initial learning rate is set to 5e-5, the number of iterations is 40, and the Batch Size is set to 16.

3) *Parameter settings in the GAT*: In order to avoid the potential problem of over-smoothing the nodes of the graph neural network, the experiment of GAT layer setting is conducted in this paper. From the experimental results, we can know that the model effect gradually decreases when the number of GAT layers is greater than two, and the model achieves the optimal effect at two layers, see Fig. 3.

4) *Graph reading methods in GAT*: There are three methods to calculate the graph representation, Sum, Mean and Max respectively, where Sum is to sum the nodes, Mean is to average the nodes and Max is to take the maximum value of the nodes.

In order to determine the optimal graph readout method and training learning rate of the model, the following experiments are set up in this paper, from the experimental results, we can see that the graph readout method is optimal

using node Sum, and the learning rate is optimal using 5e-5 for fine-tuning, see Table II and Table III.

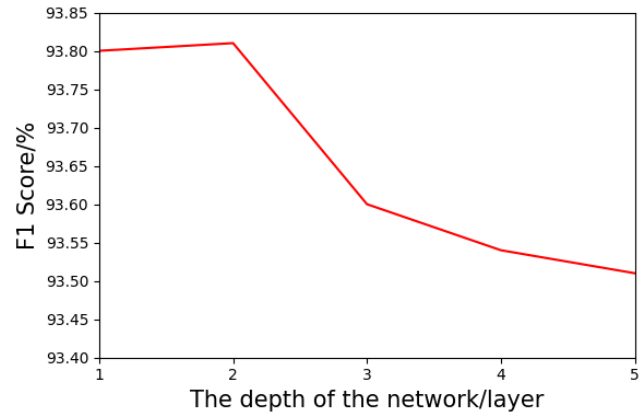


Fig. 3. Experimental results of graph attention neural networks with different layers.

TABLE II. WAYS OF READING GRAPH

Graph reading mode	Indicators			
	Acc	Precision	Recall	F1
Max	0.9379	0.9377	0.9387	0.9379
Sum	<b>0.9383</b>	<b>0.9380</b>	<b>0.9384</b>	<b>0.9381</b>
Mean	0.9352	0.9349	0.9352	0.9351

TABLE III. LEARNING RATE SETTINGS

Learning Rate	Indicators			
	Acc	Precision	Recall	F1
5e-4	0.9366	0.9362	0.9368	0.9364
5e-5	<b>0.9383</b>	<b>0.9380</b>	<b>0.9384</b>	<b>0.9381</b>
5e-6	0.9355	0.9354	0.9365	0.9355

### C. Experimental Results

CNN, BiLSTM and BertGCN are chosen as baseline models. CNN, BiLSTM and other models use word2vec trained from wiki Chinese corpus as word vector, the maximum number of single text words is 300, the dimension of word vector is 200, and different sizes of convolutional kernels are set in the convolutional layer to extract different features from different dimensions. Three sizes of convolutional kernels are set with the size of 3\* 200, 4\* 200 and 5\* 200 respectively.

The baseline experimental results are shown in Tables IV and V below, and it can be seen that the proposed method in this paper achieves the best performance in both the constructed public domain dataset and the public sentiment dataset. From the experimental results, it can be seen that the performance of the model with the introduction of graph neural network in this paper has been improved, where the biggest difference between GCN and GAT is that the way of aggregating neighbor nodes is different, and GAT utilizes attention coefficient, so the correlation between nodes calculated by GAT will be stronger to some extent. The experimental results show that the comparison of the experimental results processed using GAT is more

advantageous than GCN, which indicates that GAT is better than GCN in learning the structural information of the sentences.

TABLE IV. METAPHOR RECOGNITION RESULTS OF DIFFERENT MODELS IN DOMAIN DATASETS

Model	Indicators			
	Acc	Precision	Recall	F1
CNN	0.8839	0.8845	0.8853	0.8838
BiLSTM	0.9111	0.9109	0.9118	0.9110
Bert	0.9349	0.9346	0.9355	0.9348
BertGCN	0.9366	0.9367	0.9361	0.9364
BertGAT	<b>0.9383</b>	<b>0.9380</b>	<b>0.9384</b>	<b>0.9381</b>

The results of the experiments on the dataset ChnSentiCorp are shown in Table V. In this paper, ChnSentiCorp-6000 is selected for the experiments, and there are 3000 positive and negative sentiment sentences each in this dataset. In this public dataset experiment, model CNN, BiLSTM, AttBiLSTM, DCNN, ADCNN, and experimental result data are used from Zhu Ye et al [21], through the experiment, it can be seen that the sentiment recognition effect of BertGAT model is better than the baseline model in Precision, Recall and F1, and comparing Bert has a significant improvement, which indicates that the model proposed in this paper is advanced in emotion recognition.

TABLE V. EMOTION RECOGNITION EFFECT OF DIFFERENT MODELS IN THE PUBLIC DATASET CHNSENTICORP-6000

Model	Indicators			
	Acc	Precision	Recall	F1
CNN	0.8550	0.8557	0.8550	0.8553
BiLSTM	0.8790	0.8796	0.8790	0.8793
AttBiLSTM	0.8810	0.8827	0.8816	0.8821
DCNN	0.8865	0.8865	0.8865	0.8865
ADCNN	0.8897	0.8898	0.8895	0.8896
Bert	0.8893	0.8891	0.8897	0.8893
BertGCN	0.8850	0.8898	0.8827	0.8840
BertGAT	<b>0.9035</b>	<b>0.9045</b>	<b>0.9030</b>	<b>0.9033</b>

The main reason why the metaphor recognition model BertGAT proposed in this paper is ahead of other comparative models is the introduction of syntactic structure information and GAT graph attention network, which is based on GCN graph convolutional network and introduces the attention idea. GAT can avoid a large number of matrix operations compared with GCN, and it can also obtain the node feature information with stronger dependency.

## V. CONCLUSION

In this paper, we propose an implicit sentiment classification model BertGAT for rhetorical types. To fully capture the semantic and syntactic structural information in text sequences, the model uses the Bert pre-trained language model and the GAT graph attention network. To verify the

effectiveness of the model, experiments were conducted on both domain datasets and public datasets, and the experimental results outperformed the baseline model, which demonstrated the effectiveness of the model.

It is difficult to satisfy all downstream tasks by simple metaphor recognition classification alone. Metaphorical sentences contain mappings from ontology to metaphor, and our further research hopes to extract ontology, mapping words and their metaphors with the help of advanced Chinese named entity recognition techniques [22]. Besides, the different cognitive perspectives and cultural value orientations of English and Chinese can lead to different semantics of the same metaphorical word, and the syntactic structures of English and Chinese, two different languages, are also different [23].

we're actively engaged in follow-up this work. The next step will be to identify the similarities and differences between English metaphorical sentences and Chinese metaphorical sentences from several perspectives in order to improve the model to make it more generalizable.

## REFERENCES

- [1] Krishnakumaran S, Zhu X. Hunting elusive metaphors using lexical resources[C]//Proceedings of the Workshop on Computational approaches to Figurative Language. 2007: 13-20.
- [2] Su Chang, Fu Ze, Zheng Fakui, et al. Metaphor Recognition Method Based on Dynamic Classification [J]. Journal of Software, 2019, 30(11): 3340-3354.
- [3] Su Chang, Wang Xiaomei, Huang Shuman, Chen Yijiang. An Approach to Metaphor Comprehension Based on Relevance Constraints [J]. Journal of Software, 2017, 28(12): 3167-3182.
- [4] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.
- [5] Li Bin, YU Lili, Shi Min, et al. Simile Calculation of Image [J]. Journal of Chinese Information Processing, 2008, 22(6): 27-32.
- [6] Shutova E, Sun L, Korhonen A. Metaphor identification using verb and noun clustering[C]//Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010). 2010: 1002-1010.
- [7] Zhu Jiaying, Wang Rongbo, Huang Xiaoxi, et al. Multilevel Metaphor Recognition Method Based on Bi-LSTM [J]. Journal of Dalian University of Technology/Dalian Ligong Daxue Xuebao, 2020, 60(2): 209-215.
- [8] Su C, Fukumoto F, Huang X, et al. DeepMet: A reading comprehension paradigm for token-level metaphor detection[C]//Proceedings of the second workshop on figurative language processing. 2020: 30-39.
- [9] Su Kuilin, Zhang Kai, Lv Xueqiang, Zhang Le. Noun Metaphor Recognition Based on Fusion Model [J]. Computer Technology and Development, 2022, 32(06): 192-197.
- [10] Zhang S, Liu Y, Ma Y. SaGE: Syntax-aware GCN with ELECTRA for Chinese Metaphor Detection[C]//Proceedings of the 20th Chinese National Conference on Computational Linguistics. 2021: 667-677.
- [11] EKATERINA S, SIMONE T, ANNA K. Statistical metaphor processing[J]. Computational Lingus, 2013, 39 (2) : 301-353.
- [12] Kenton J D M W C, Toutanova L K. Bert: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of naacL-HLT. 2019: 4171-4186.
- [13] Zhang Dongyu, Cui Zijuan, Li Yingxia, Zhang Wei, Lin Hongfei. Noun Metaphor Recognition Based on Transformer and BERT [J]. Data Analysis and Knowledge Discovery, 2020, 4(04): 100-108.
- [14] Tian Y , Chen G , Song Y , et al. Dependency-driven Relation Extraction with Attentive Graph Convolutional Networks[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference

- on Natural Language Processing (Volume 1: Long Papers). Online: Association for Computational Linguistics. 2021: 4458-4471.
- [15] Caren Han S, Yuan Z, Wang K, et al. Understanding Graph Convolutional Networks for Text Classification[J]. arXiv e-prints, 2022: arXiv: 2203.16060.
- [16] Wu Bo, Liang Xun, Zhang Shusen, et al. Advances and Applications of Graph Neural Networks [J]. Chinese Journal of Computing, 2022, 45(1).
- [17] Ji T, Wu Y, Lan M. Graph-based dependency parsing with graph neural networks[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 2475-2485.
- [18] Tian Y, Chen G, Song Y. Enhancing aspect-level sentiment analysis with word dependencies[C]//Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. 2021: 3726-3739.
- [19] Fu Chaoyan, Huang Xianying, Liu Hankai, Qi Songzhe. Syntactic feature fusion enhancement aspect level sentiment analysis [J/OL]. Journal of Chinese Mini-Micro Computer Systems:1-10[2022-10-05].<http://kns.cnki.net/kcms/detail/21.1106.TP.20220215.1046.021.html>
- [20] TAN S, ZHANG J. An empirical study of sentiment analysis for Chinese documents[J]. Expert Systems with Applications, 2008, 34(4): 2622-2629.
- [21] Zhu Ye, Chen Shiping. A Novel approach to Emotion Analysis Based on Convolutional Neural Networks [J]. Journal of Chinese Mini-Micro Computer Systems,2020,41(03):551-557.
- [22] Deng Yiyi, Wu Changxing, Wei Yongfeng, Wan Zhongbao, Huang Zhaohua. A Review of named Entity Recognition Based on Deep Learning [J]. Journal of Chinese Information Processing,2021,35(09):30-45.
- [23] Liu Xiaoyu, Liu Yongbing. A Comparative Study of "Big/small" Spatial Metaphors in English and Chinese from the Perspective of Cognition [J]. Journal of PLA University of Foreign Languages ,2022,45(02):17-24.



# User Perceive Realism of Machine Learning-based Drone Dynamic Simulator

Damitha Sandaruwan<sup>1</sup>, Nihal Kodikara<sup>2</sup>, Piyumi Radeeshani<sup>3</sup>,  
K.T.Y. Mahima<sup>4</sup>, Chathura Suduwella<sup>5</sup>, Sachintha Pitigala<sup>6</sup> and Mangalika Jayasundara<sup>7</sup>  
University of Colombo School of Computing,  
35 Reid Ave, Colombo 00700, Sri Lanka<sup>1,2,3,4,5</sup>  
Department of Statistics Computer Science, University of Kelaniya  
Sri Lanka<sup>6,7</sup>

**Abstract**—The drone will be a commonly used technology by a significant portion of society, and simulating a given drone dynamic will be an essential requirement. There are drone dynamic simulation models to simulate popular commercial drones. In addition, there are many Newtonian and fluid dynamics-based generic drone dynamic models. However, these models consist of many model parameters, and it is impracticable to evaluate the required model parameters to simulate a custom-made drone. A simple method to develop a machine learning-based dynamic drone simulation model to simulate custom-made drones mitigates the issues mentioned above. Specifically, the authors' research is associated with the development of a machine learning-based drone dynamic model integrated with a virtual reality environment and validation of the user-perceived physical and behavioural realism of the entire solution. A figure of eight manoeuvring patterns was used to collect the data related to drone behaviour and drone pilot inputs. A Neural Network-based approach was employed to develop the machine learning-based drone dynamic model. Validations were done against real-world drone manoeuvres and user tests. Validation results show that the simulations provided by machine learning are accurate at the beginning and it decreases the accuracy with time. However, users also make mistakes/misjudgments while perceiving the real-world or virtual world. Hence, we explored the user perceive motion prediction accuracy of the simulation environment which is associated with the behavioural realism of the simulation environment. User tests show that the entire simulation environment maintains substantial physical realism.

**Keywords**—Drone; simulation; machine learning; drone dynamics; virtual reality

## I. INTRODUCTION

Drone is an unmanned aircraft controlled by a ground base station, used primarily to carry out air-based missions, such as surveillance, transportation and entertainment [1], [2]. Historically drones were first used by the military for ground surveying and spying missions, and they were called Unmanned Aerial Vehicles (UAVs) [3], [4].

Nowadays the term drone is widely used for quadcopters, vertical takeoff and land (VTOL) type miniaturized aircraft, with four propeller blades, pushing air downwards to maintain itself on air, while moving back and forward using the thrust created by the same propellers [5]. Such drones come in many sizes ranging from a few inches up to several feet and are used in many applications including for toys, entertainment, photography, land surveillance and scientific research [1], [6], [7], [8].

According to the Stanford University Intelligent Systems Laboratory and National Aeronautics and Space Administration (NASA), within the next few years, Low-Altitude airspace will be congested with Low-Altitude Remotely Piloted Aircrafts [9]. This will be a common technology used by a significant portion of society. Due to the maturing of the technology, it requires a skilled and qualified human resource to use the technology and train pilots for specific tasks and missions. At the same time, it allows individuals and various organizations to build custom-made drones at their own pace.

Therefore, drone pilot training for custom-made drones via simulations is essential. It provides the knowledge and practical skills that are necessary to safely and efficiently operate unmanned aircraft for commercial and non-commercial use [10], [11]. Hence, drone pilot training simulators will be a very important requirement. Depending on the safety-critical level of the drone operation, there is a requirement of simulates the drone operations prior to the real real-world operation. This type of simulation supports the identification of potential disasters due to the manoeuvring capability of the drone and the pilot. Hence, drone simulators with required realism will be essential in future [1].

There are solutions to simulate real-time drone dynamics, such as the DJI drone simulator. These simulators are capable of simulating the bundled series of drones belonging to a particular vendor. There are many proposed generalized drone dynamic simulation models and most of these models are based on Newtonian dynamics and fluid dynamics. Simulating custom-built drones with these generalized drone dynamics is a highly challenging task. It requires evaluating model parameters related to the custom build drone and it needs many experiments to be conducted with the given ideal conditions, domain-specific knowledge and a wide range of practical issues. A machine learning-based drone dynamic model avoids most of the above issues. There is significant value to explore the possibility of developing a machine learning-based drone dynamic model and incorporating it with a Virtual Reality (VR) environment to simulate a custom-built drone. It is highly essential to evaluate the realism level of such a drone dynamic model and the VR environment. Overall realism of a virtual environment can be expressed with physical realism and behavioural realism. Behavioural realism expresses the accuracy of dynamic activities such as motion predictions. Physical realism expresses the physical infrastructure of the simulated environment [12], [13], [14], [15].

The remainder of this paper is structured as follows: State-of-the-art drone simulation models are critically evaluated in Section II. Section III discusses initial research initiatives such as data collection techniques. The development processes of the machine learning model and VR environment for the proposed dynamic drone simulator are covered in Sections IV and V, respectively. Section VI presents the experimental and validation results of the simulator under different criteria. Finally, Section VII, concludes the paper along with prospective research directions.

## II. RELATED WORKS

There are many proposed drone dynamic simulation models and most of these models are based on Newtonian dynamics and fluid dynamics [16], [17]. These models were used to simulate drones for different applications and comparisons of different simulation models were reported [18], [19]. However, the evaluation of required model parameters and simulations of existing drones is not discussed.

There are commercial simulators that could use for the simulation of drones. For example, the DJI Assistant 2 software is a drone simulator provided by DJI used to simulate selected DJI drones [20], [21]. The DJI Assistant 2 is programmed/programmable to simulate drones with offline remote control data. Real Flight drone/flight simulator [22], Simpro drone simulator [23], Liftoff by Immersion RC [24] and HELIX professional R/C flight simulator [25] are some of the reviewed commercial drone simulators which can be employed to simulate particular commercial drone.

There are many research and development works carried out by employing machine-related theories in drone-related ICT solution developments [26], [27], [28], [29], [30], [31], [32]. However, most of these applications are related to designing and developing autonomous drones and target tracking in outdoor/indoor environments.

Jemin et al. presented a method to control a quadrotor with a neural network trained using reinforcement learning techniques [33]. They demonstrated the performance of the trained policy both in simulation and with a real quadrotor. The trained policy shows outstanding performance and remains computationally cheap simultaneously. Furthermore, it shows many other advantages of neural network policies that are not limited to their versatility. This experiment is limited to a small space covering approximately a 2m x 2m x 2m controlled area.

Osman Çakira and Tolga Yükselb [34] developed a neural network-based controller for quadrotors. In this study, neural network control of quadrotors is aimed to obtain an artificial intelligence-based drone controller and the results show that neural network controllers achieve satisfactory trajectory tracking results. This experiment is also limited to a small space that approximately covers 5m x 3m x 3m controlled area.

Jeong, Baek and Lee propose a prediction model of the vehicle trajectory [35]. Their approach is not based on physics-based motion models and there are no kinematic and dynamic models, laws of physics and fluid dynamics. They employed a Deep Neural Network that takes as input vehicle velocity, acceleration, yaw rate, steering, and road curvature. The authors

discussed the advantages of employing deep neural network-based predictions to avoid several potential issues in similar physics-based vehicle simulation requirements.

Jeong et al. proposed Deep Neural Network (DNN) that considers preprocessed vehicle velocity, acceleration, yaw rate, steering, and road curvature as the input layer and eventually reaches the output layer via multiple hidden layers. DNN uses an activation function with a function called a rectified linear unit (ReLU) [36], [37]. However, a ReLU function is not a perfect match to be used in this study because the final outputs of our DNN model include expected future lateral movements, which can be negative. Hence, Leaky ReLU was employed for the activation function, which is slightly different from the original ReLU [35]. The results of the proposed study confirm the feasibility of Deep Neural Network-based long-term trajectory prediction for vehicles driving on roads with a certain level of road conditions such as varying curvature.

Jackson et al. employed the machine-learning technique to design and develop a dynamic model of a rotorcraft [38]. Their key justification for this approach is that widely used physical-law-based and substantially accurate rotorcraft dynamic models need to be simplified to make real-time motion predictions. Hence, it leads to motion prediction errors compared to the real vehicle. In their current work, machine-learning techniques are employed to train a rotorcraft dynamic model to predict the dynamic on-axis motion responses such as pitch rate, roll rate and yaw. The employed machine learning was designed with a Gaussian Process (GP) non-linear autoregressive model [39]. They have proven that the machine-learning approach can be successfully utilized to predict the on-axis motions of a rotorcraft. The obtained level of accuracy is generally higher than the physics-based dynamic models.

Punjani proposed a helicopter dynamic modelling method with a Rectified Linear Unit (ReLU) Network Model [40]. The reasons for selecting this approach are helicopter has a complex dynamic system with rigid body dynamics with aerodynamics, engine dynamics, vibration and other factors such as manoeuvring patterns. They described several baseline models and shows that the helicopter dynamics with ReLU significantly outperformed other considered baseline models. Furthermore, It improves acceleration prediction over state-of-the-art methods and they presented performance gains techniques with hyperparameters fine-tuning. They selected a range of manoeuvres such as forward/sideways flight, vertical sweeps, inverted vertical sweeps, stop-and-go, flips, loops, turns, circles, dodging, orientation sweeps, orientation sweeps with motion, gentle freestyle and aggressive freestyle.

Considered, baseline models are based on Linear Acceleration Model and it serves as a direct state-of-the-art performance baseline. Validations, efficacy investigations, compare and contrast among baseline models and the ReLU-based model were done by using data obtained from the Stanford Autonomous Helicopter Project [41]. Following Fig. 2 presents observed and predicted accelerations in the up-down direction for selected three different aerobatic manoeuvres. It shows that the baseline Linear Acceleration model performs poorly compared to the novel ReLU Network Model [40].

Sandaruwan et al. proposed a machine learning-based approach to simulate drone dynamics related to the figure of

Eight Manoeuvring pattern [1]. Authors have obtained satisfactory results and proven that the machine-learning approach can be successfully utilized to predict drone motions. However, they have not published the evaluation of the developed drone dynamic model and simulation environment.

Zhang et al. investigated users' situational awareness of virtual outdoor, virtual indoor, real-world indoor, and real-world outdoor environments. They considered distance judgment for their experiments and identified potential error factors and other considerations [42]. They also investigate the impact of users' real-world scene awareness on distance judgment in the same simulated scene in a virtual environment. Their results suggest that both the virtual and real-world environments have an impact on distance judgment in VR which affects the users' ability to perceive the realism of VR environment [42].

Ziemer et al. also explored the order in which people experience real and virtual environments that influence their distance estimates. They also identified potential error factors, error percentages and other considerations in estimating distance in real and virtual environments. Their results provide facts and figures to measure users' situational awareness of the virtual and real scenes [43]. They presented the importance of presence and reality judgment in the VR environment and aspects that contribute to creating a person's reality judgment in a given scenario in the VR environment. They discuss the problem of "How do people decide whether something is real or not?". Their work aims to design a self-report measure that assesses both constructs [44].

Sandaruwan et al. conducted several research studies to design and deploy a Maritime VR Environment [14], [13]. They discuss essential factors/considerations related to validating the user perception of the physical & behavioural realism of a maritime VR environment. They employed several techniques & methods to validate the physical & behavioural realism of the maritime VR environment. Moreover, they used techniques such as comparing real vehicle trials and simulated vehicle trials, simulation of simple possible scenarios and user tests [15]. However, those techniques & methods are applicable to most VR environments with dynamic vehicle models.

By considering the above-reviewed literature and the main objective of designing and developing a machine learning-based drone dynamic model for outdoor simulation of a custom-built drone, the following actions were executed: 1) Deep Neural Network-based approach with Leaky ReLU or any other appropriate activation function. 2) The figure of eight manoeuvring pattern-based outdoor data collection to build the machine learning model. 3) Conduct evaluation/validation such as comparing real drone trials and simulated drone trials, simulation of simple possible scenarios and user tests to measure the user perceive realism of the develop machine learning-based drone dynamic model.

Analysis of Newtonian dynamic and fluid dynamic-based analytical solution used by the researchers [16], [17], [18], [19] shows that takeoff & landing of the drone is more complicated and needs more accurate complex dynamics. In addition, real-world wind flows with a turbulent effect make the situation further complicated. Hence, this research does not focus on the takeoff & landing of the drone, and all experiments and tests were carried out in a calm outdoor environment with

negligible wind effect.

### III. PRELIMINARIES AND DATA COLLECTION

Four-rotor drones are a subset of multirotor systems and these drones use four rotors to keep them flying. A popular example of these multirotor drones is the widely used Phantom drone made by the SZ DJI Technology Co. Ltd [45]. Drone movements are controlled by a handheld radio controller. The drone changes its position and orientation based on the given radio controller inputs. This research focuses on a machine learning-based four-rotor drone dynamic model to build the relationship between the drone pilot's radio controller inputs and the drone's position and orientation. The Phantom drone by the SZ DJI Technology Co Ltd was selected for the experiments and it has six degrees of freedom motions (Three rotational motions & three translational motions).

DJI Phantom quadcopter drone comes with four propulsors that enable vertical takeoff and landing. It has four key controllable variables which move the drone in the 3D space [45], named throttle, pitch, roll & yaw. Fig. 1 illustrates a radio controller with key controllable variables. Based on the key controllable variable inputs, the drone changes its position and orientation.

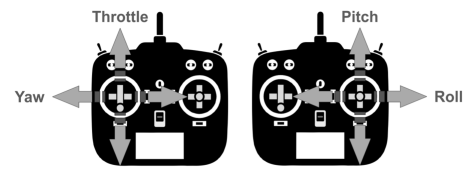


Fig. 1. A radio controller with key controllable variables.

Movement in the horizontal frame is achieved by tilting the platform with the different thrusts of the motors. Vertical movement is achieved by changing the total thrust of the motors.

DJI Phantom 4 [45] consists of many inbuilt sensors such as accelerometers, proximity sensors, GPS and GLONASS. That sensor assists with precision flying, precise hovering and much more. In addition, it provides a flight log consisting of flight position and orientation data & radio controller input data with the frequency of 10 sample points per second (10Hz). DJI Phantom 4 Pro [46] drone maintains 1.5-meter position accuracy with inbuilt GPS sensors & it can be enhanced up to 1cm accuracy by configuring external sensors.

The required machine learning-based drone dynamic model must predict the drone's position and orientation against the radio controller input. Hence, the required data can be categorized into two main categories as given below:

- Drone pilot inputs are entered via the drone Radio Controller (RC) and represent the given inputs, such as throttle and rudder values that are responsible for the drone position and orientation changes.
- The Drone's position and orientation change with time, and all other relevant sensor information vary with time (E.g. Battery level, accelerations, velocities).

If all manoeuvres are performed with fully charged batteries and the battery level is within 97% and 100%. Then it reduced the potential performance variation of the drone due to the battery power variations. The recorded data set consists of over forty parameters that vary with time. After examining the time-varying fields of the raw data set, the following observations were made:

- Certain sets of parameters directly imply the other set of available parameters. E.g. Latitude, Longitude and Altitude present the position of the drone in 3D space. All data stamps were recorded with a constant time gap. VelocityX, VelocityY and VelocityZ also present the position of the drone in 3D space.
- There are certain sets of parameters that are constant during the experiment or indicate negligible variation. E.g. GpsCount, GpsLevel Battery Power(%), Battery Voltage, Battery Voltage Deviation, Battery Cell Voltages.
- There are other sets of parameters that are static due to the selected settings of the drone or provide Tips and warnings. E.g. App Tip, App Warning, App Message and Flight Mode.

Hence, after considering the above fact extracting the required data and selected key parameters of the data set are described below:

- Time (seconds): Time elapsed since the power-up.
- RcAileron: Rc signals for roll.
- RcElevator: Rc signals to control the horizontal pitch attitude of the drone.
- RcRudder: Rc signals to control the yaw of the drone.
- RcThrottle: Rc signals control the engine's speed and indicate how fast or slow the drone's movement.
- x,y,z: Cartesian conversion of Longitude and Latitude.
- Orientation: Bearing of the head of the drone in degrees.

The above parameters describe the drone pilot input via the radio controller & resultant position and orientation of the drone. Hence, the above parameters were taken into consideration to develop a machine-learning drone dynamic model.

#### IV. DEVELOPMENT OF MACHINE LEARNING MODEL

This research focuses on a machine learning-based drone dynamic model based on a subset of artificial intelligence. Machine learning consists of several subsections. Deep learning is one of the subsets of machine learning in which artificial neural networks adapt and learn from a large amount of data [47].

Machine learning models/methods or learnings are based on what it has learned only. Neural network structures/arrange algorithms in layers of fashion that can learn and make intelligent decisions on their own [48]. Neural networks are more suited to solve complex machine-learning problems. Neural networks can learn and model the relationships between inputs and outputs that are nonlinear and complex. This can be

used to generalize input-output relationships, and reveal hidden relationships, patterns and predictions. It supports modelling highly volatile time series data and capable of predictions [49]. Neural networks require much more data than traditional Machine Learning algorithms to complete the model development. Depending on the requirement, building a customized neural network model that is perfectly suited. However, it takes more time compared to the traditional ML algorithm. A neural network consumes a longer time to train rather than a traditional machine learning model. It requires continuous computational resources, depending on the architecture of the neural network and the size of the data [50].

This research deals with complex rapidly changing input data set and output data set (Drone pilot's radio controller inputs and drone's position and orientation). The relationships between inputs and outputs are nonlinear. A single drone pilot trial consists of thousands of data points and the entire data set consists of over 200 thousand data points. A continuous computational resource is not a vital issue with the available technological infrastructure. Hence, an Artificial Neural Network (ANN) based approach was selected.

An artificial neural network has parameters that cannot be directly estimated from the data. This type of model parameter is referred hyperparameter. No analytical solution is available to calculate appropriate values for hyperparameters [51]. An artificial neural network has many hyperparameters. However, two key hyperparameters are the number of layers and nodes in each hidden layer. It controls the entire architecture/topology of the artificial neural network. In addition, there are other hyperparameters such as activation function, the number of epochs, batch size, learning rate, Mini-batch size, and Learning rate, which are identified as other potential hyperparameters [52].

Several frameworks and libraries have been developed in the last few years to fulfil machine learning-related necessities. Industry and academia use various frameworks and libraries to expedite the neural network-based model development, training and good results. Hence, model development and training have become easier. Based on the star ratings on Github, and our similar project experiences in the field, TensorFlow [53] was selected as the most effective and easy-to-use framework and library.

TensorFlow is a full-fledged open-source deep learning framework designed and developed by Google. It was initially released in 2015 and it comes with documentation, training support, scalability options and support for different platforms. In addition, TensorFlow is associated with flexible, comprehensive community resources, libraries, frameworks and tools that facilitate developing and deploying machine learning solutions [53]. Keras [54] is a high-level neural network library that runs on top of TensorFlow. Further, Keras supports building high-level API to be used for easily building and training models. Keras is a built-in Python. Keras is an open-source software framework that provides a Python interface for designing and developing artificial neural networks.

On top of TensorFlow and Keras, a sequential machine-learning model was developed and tuned to produce optimum results. The developed model has three layers, consecutively 56 and 112 nodes in the first and second hidden layers and one

node in the output layer. The activation function for all three layers is linear. The loss is calculated using the mean squared logarithm error while using the Adam optimizer.

### V. DEVELOPMENT OF VR ENVIRONMENT

According to the 3D graphics rendering pipeline and commonly used game engine architectures, this type of 3D drone simulation environment consists of a real-time computational drone motion prediction module, configuration module (Configure environment, drone drill, etc.), The visual rendering engine, sound generation engine and seamless display system/head-mounted display [55]. Real-time motion prediction carries out by the proposed machine learning model which considers user interactions and defined environmental conditions to predict real-time drone motions. The visual rendering engine considers the predicted state of the drone-requested view of the virtual environment (first-person view, third-person view, etc.) to generate the relevant scenery. Finally, the user can see the generated visual through a seamless display system or head-mounted display. Integration of a sound generation engine brings more realism to the solution. A high-level structure of the virtual environment is illustrated in Fig. 2.

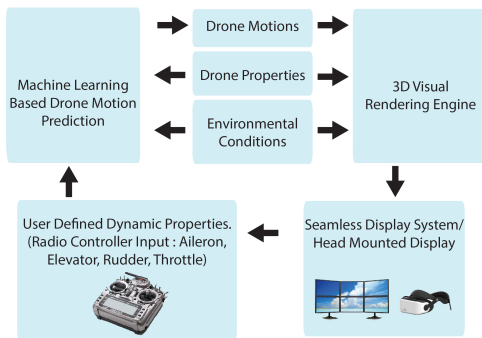


Fig. 2. The high level structure of the 3D virtual environment

There are many open-source, free and commercial frameworks, libraries and engines to develop 3D VR environments [56], [57], [58], [59]. Mairaj, et al. carried out comprehensive literature review related to the design and development of drone simulators and analyzed available frameworks, libraries and engines [60]. Based on the Mairaj, et al. review, authors' own experience gained during the last ten years [61], GitHub Star rating and other reviews, Microsoft Aerial Informatics and Robotics Simulation (AirSim) [62], [63] open-source robotics simulation platform was selected to design and develop the proposed 3D virtual environment to simulate drone dynamics.

AirSim is developed for AI research to experiment with deep learning, computer vision and reinforcement learning algorithms for autonomous vehicles. However, AirSim provides APIs to retrieve data and control vehicles in a platform-independent way. Hence, several modules of AirSim were decoupled and several API facilities, such as retrieve data and control vehicles were slightly modified to develop the proposed 3D drone simulation environment [62]. Fig. 3 presents the high-level architecture of the modified Microsoft AirSim

simulation platform which was used to develop the 3D drone simulation environment.

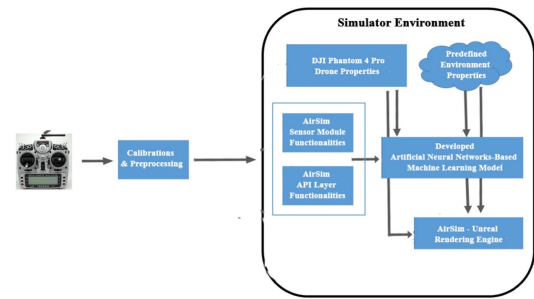


Fig. 3. High-level architecture of the modified microsoft airsims simulation platform

AirSim simulation platform was configured with an Unreal environment because the Unreal Marketplace has several environments available that can be used to generate realistic 3D scenarios easily. As illustrated in Fig. 3, AirSim physics engine was replaced with a developed machine learning-based drone dynamic model. The selected functionalities of the AirSim sensor module and AirSim API layer were reused to connect the real physical radio controller and virtual environment. However, a real Fr-Sky Taranis radio transmitter & receiver was used to capture drone pilot inputs and feed input stream to develop a machine learning-based drone dynamic model in the simulated environment. Hence, a separate calibration/preprocessing module was developed to align/map Fr-Sky Taranis radio transmitter inputs and DJI Phantom 4 Pro drone radio controller inputs. Fig. 4 presents the Fr-Sky Taranis radio transmitter/radio controller, which can be configured to align with DJI Phantom 4 Pro drone radio controller.



Fig. 4. Fr-sky taranis radio transmitter/radio controller

Fig. 5 presents the design and development of the proposed machine learning-based drone dynamic simulation environment. It uses an Intel Core i7 CPU with 3.6 Hz, GTX 1070 graphic card and Ram 32 GB to run the entire solution-inducing Unreal rendering engine and machine learning model. 32-inch screen connected to visualize the rendered image stream, and it makes continuous sensation for the users (Drone pilots) who interact with the virtual environment via Fr-Sky Taranis radio controller. Depending on the user's/drone pilot's preference instated of the screen, a head-mounted display such as Oculus Quest-2 can be connected with the solution.

### VI. EXPERIMENTAL RESULTS AND VALIDATION

The entire simulated environment can be validated under different criteria. As discussed before, simulation represents



Fig. 5. Developed machine learning based drone dynamic simulation environment.

a real-world scenario with assumptions, limitations and simplifications. Comparison of the simulated results against its actual real-world scenario is one of the validation methods employed in this type of research and development work. This can be done by considering comparisons such as positions and orientations with the timestamps. The user test is another vital validation technique to validate the user's perception of the simulated environment. In this technique, the virtual environment is validated based on the observations of pilots with much drone experience.

Some of these validation techniques are quantitative, while others are qualitative. For example, quantitative validation methods can be used to validate activities such as the accuracy of the motion predictions and qualitative validation methods can be used to validate components such as user perception enhancement and ecological validity of the simulated environment.

The validation process of the proposed virtual environment was divided into the following four segments:

- Carry out short-term motion predictions with the developed machine learning model: In this approach, real-world scenarios were simulated in the simulated environment and investigate the accuracy of the predictions against the radio controller input variations.
- Carry out long-term predictions with simple possible scenarios and investigate the acceptability of the obtained results.
- Carry out long-term motion predictions: In this approach, real-world scenarios were simulated in the simulated environment and compare-contrast the predictions against the real-world scenario.
- User tests: Compare the real-world user perception and the user perception of the VR solution with drone pilots who have much experience with drones. This can be used to compare the numerical and user-perceived accuracy and ecological validity of the VR solution.

As mentioned in the related work, we analyzed previously carried out research work such as [13], [14], [15], [42], [43], [44]. The analysis shows that users' awareness of the location of a real-world outdoor scene or virtual-world outdoor scene consists of percentage-based errors ranging from 10% to 25%.

As motioned above, the developed machine learning model predicts the drone's position and orations against the radio controller inputs throttle, pitch, roll & yaw. Due to the assumptions and limitations, it works under negligible wind effects while drones perform forward and lateral movements. Under the simulation of short-term motion predictions, a single prediction was made by considering the drone's current position, oration and controller inputs throttle, pitch, roll & yaw. We considered several figures of eight shape drone trials and considered the drone's actual position, orientation and ratio controller inputs, then predicted the drone's position and orientation after 0.1 seconds. Then closely investigate the effect of the factors on prediction error. Fig. 6 illustrates the Actual positions of the drone and predicted positions of the drone.

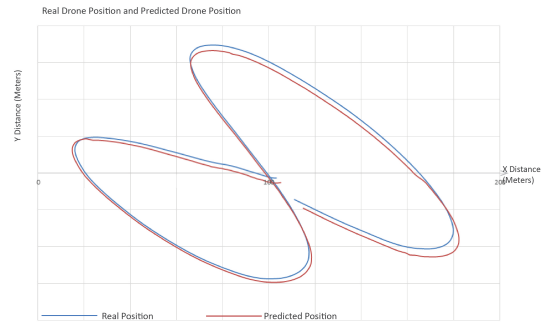


Fig. 6. Actual positions of the drone and predicted positions of the drone.

Fig. 7 illustrates the variation of the prediction error of the developed machine learning model against the time. It shows that the variation of the error is within 0.5 meters, and it is required to investigate the factors that affect the error.

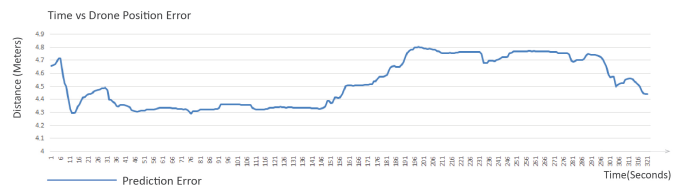


Fig. 7. Variation of the prediction error of the machine learning model against the time.

Table I presents the correlation between radio controller inputs (RcAileron, RcElevator, RcRudder, RcThrottle) and the machine learning model's prediction error. Further, it shows that the prediction error highly depends on the RcRudder-radio controller input.

Fig. 8 illustrates the acceleration error of the machine learning model against time. Fig. 9 depicts the magnified segment of Fig. 8 that illustrates the acceleration error of the machine learning model against time.

Fig. 8 and 9 show that most of the time, acceleration error is significantly less. Compared with the rudder variation pattern and acceleration error, it shows that the acceleration error rapidly increases with the rudder. However, smaller rudder variations in the developed machine-learning model

TABLE I. CORRELATION BETWEEN CONTROLLER INPUTS (RCAILERON, RCELEVATOR, RCRUDDER, RCTHROTTLE) AND PREDICTION ERROR

Correlation between Prediction Error and Radio Controller Inputs	Correlation Coefficient
Correlation between Prediction Error and RcAileron	0.003580433
Correlation between Prediction Error and RcElevator	0.369202804
Correlation between Prediction Error and RcRudder	0.989092821
Correlation between Prediction Error and RcThrottle	0.242975470

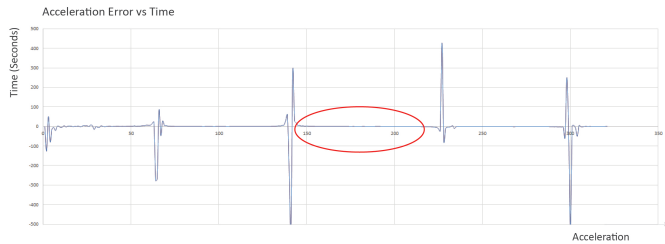


Fig. 8. Acceleration error of the machine learning model against time.

provide substantial accuracy. As previously mentioned, a real-world outdoor scene or virtual-world outdoor scene consists of percentage-based errors ranging from 10% to 25%. Hence, the developed model and its short-term motion prediction results provide encouraging results to perform long-term predictions and further validate the model.

#### A. Long-Term Drone Motion Predictions with Simple Possible Scenarios

As explained above (Subsection: Design and development of the VR solution), the Fr-Sky Taranis radio transmitter & receiver was connected to the solution and make enabled user interactions via an actual radio controller. Fig. 5, presents the designed and developed machine learning-based drone dynamic simulation environment. Three experienced pilots were exposed to the VR environment and asked to perform simple drills such as the figure eight type manoeuvre, manoeuvre along a straight line and circular manoeuvre. Their responses were assigned to a typical Likert scale with a five-point agreement scale related to simulation results produced by the solution. The five points of the Likert scale are strongly agreed, agree but no idea, disagree, and strongly disagree. The overall Likert scale results show that the three experienced pilots agreed with the results produced by the designed and developed simulation environment.

#### B. Long-Term Drone Motion Predictions and Validate against Real-World Scenario

Under the assumptions and limitations of the drone dynamic model, a simulation of long-term motion predictions was made. We consider several figures of eight-shape drone trials for this validation. We considered the drone's initial settled/stabled position, orientation in the outdoor environment and ratio controller inputs, then predicted the drone's position, and orientation continuously. Then closely investigate the deviations between the actual drone trial results and predicted results. Fig. 10 and Fig. 11 illustrate the actual trajectory of the drone and the predicted trajectory of the drone.

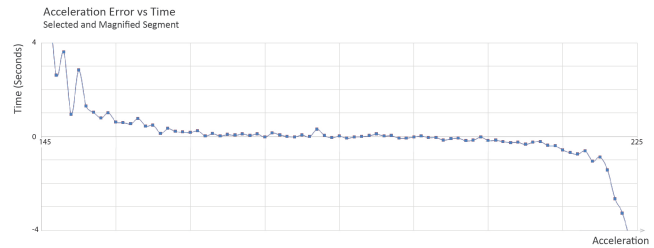


Fig. 9. Magnified segment of the previous graph that presents the acceleration error against time.

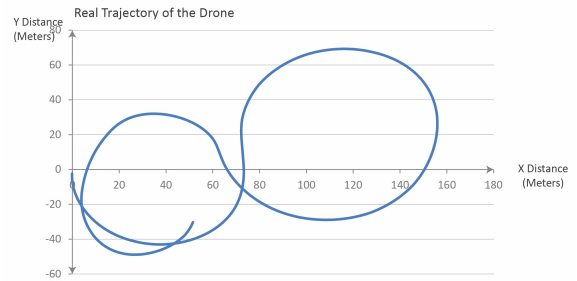


Fig. 10. Actual trajectory of the drone.

The above results show that the predictions are deviated/deviating with the time, and initial position predictions are closer to the actual position of the drone. As we discussed above, users perceive the position of an outdoor drone or users perceive the position of a drone in an outdoor VR environment is accolated with a percentage-based error ranging from 10% to 25%. We consider the average of this error and define error merging for the predicted results. Fig. 12 illustrates the initial segment of the predicted drone trajectories and real drone trajectory with probable user perceive rejoin of the drone.

#### C. User Tests

The developed machine learning-based drone dynamic simulation environment needs to be validated to determine the immersive feeling "sense of being there" or "how users perceive" in the VR environment. The most common method of measuring this presence or "sense of being there" is to use questionnaires [64], [65]. Questionnaires give subjective measurement, and in most questionnaires, participants' responses to each question are assigned to a numerical scale [14]. Finally, the immersive feeling "sense of being there" in the VR solution can be reflected as a percentage. In the research validation phase, our primary focus is on subjective measurements. Hence, the user test was designed with a questionnaire. It targets three experienced drone pilots, including the drone pilot involved in the data gathering/recording pace, and the user test can be summarized as follows:

Simulate known conditions and record experienced drone pilots' responses. First, concerning each participant's response (feeling about the simulated scenario), qualitative properties of the simulated scenario will be assigned to a numerical scale (Likert scale) as follows [66]. Next, the quantitative properties of the simulated scenario will be directly recorded. Finally, the deviation from the expected value will be calculated. Under

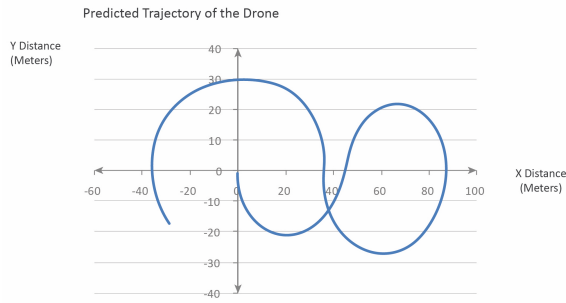


Fig. 11. Predicted trajectory of the drone

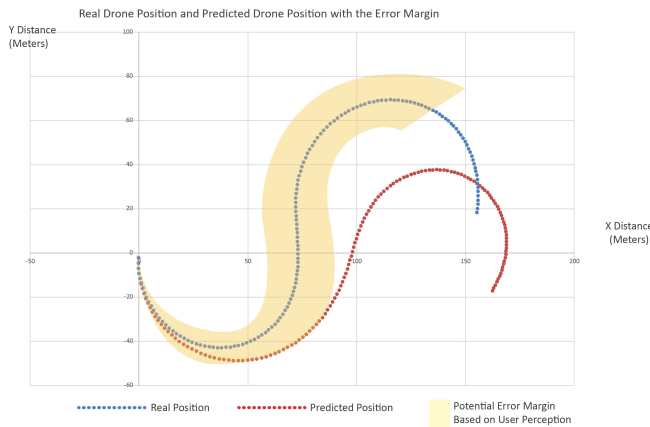


Fig. 12. Initial segment of the predicted drone trajectory and real drone trajectory with probable user perceive rejoin of the drone.

the simulated scenarios, many responses were recorded and selected questions/statements are given below:

- Approximated travel distance  $\approx 300\text{m}-400\text{m}$   $\rightarrow$  What is the approximated travel distance?
- What is the current velocity?
- Asked to perform figure eight manoeuvre. What is the approximate size of the performed figure eight manoeuvre?
- Compared to the real-world figure eight manoeuvre with Phantom 4 Pro, the drones' response to the radio controller during that figure eight manoeuvre carried out in the simulated environment is realistic.
- Used radio controller: Ease of Use/ realism level compared to the real-world Phantom 4 Pro radio controller is it perfect/realistic?
- How many vehicles were there on the ground surface?
- Presence (the user's sense of immersion or "being within" the environment) is perfect.
- Three drone pilots were exposed virtual environment with large flat screen-based visualization. Do you recommend HMD or any other visualization method?

The actual and simulated scenarios' qualitative and quantitative properties were compared throughout this user test.

TABLE II. SELECTED RESULTS OF THE CARRIED OUT USER TESTS

Simulated Scenario	Pilot Response Numerical Indicator
Physical realism of the entire VR environment (Input Methods/ Controllers/Output methods/ Visualizations)	1.4 (from -2 to 2)
Realism of the Drone Maneuverability	1.0 ( from -2 to 2)
Deviation of the size of the performed trial/covered area during the circular/ figure eight maneuvers	20%
Deviation of travel distance judgment	25%
Deviation of speed judgment	15%
Awareness and information gathering capability	1.4 (from -2 to 2)
Presence (immersion or "being within" the environment)	1.0 (from -2 to 2)

The main focus is to identify spatial awareness (The user's implicit knowledge of his position and orientation within the environment - during and after travel), information gathering (the user's ability to actively obtain information from the environment - during travel), and accuracy of the drone motion prediction against the real-world situation. Table II presents selected results of the carried out user tests.

## VII. CONCLUSIONS AND FUTURE WORK

The overall objective of the presented work is to propose a machine learning-based drone dynamic model and VR environment that can simulate existing drones without domain-specific knowledge and sophisticated laboratory infrastructure. Under selected circumstances, the proposed solution's accuracy and user-perceived accuracy were evaluated, and the authors were able to get promising results. The following conclusions and recommendations can be made based on the entire research.

- The proposed and developed machine learning model and its evaluation experiments were carried out by using Commodity-Off-The-Shelf hardware.
- If an accurate location tracker is available, the same procedure can be followed, and a similar machine-learning model can be developed for any existing drone.
- Short-term predictions of the proposed and developed machine learning model are within the user-perceived accuracy of both the real-world outer door scene and the virtual world outer door scene.
- The accuracy of the predictions mainly depends on the rudder variation. If the rate change of the rudder is more significant, then predictions of the proposed and developed machine learning model are less.
- Long-term predictions of the proposed and developed machine learning model are within the user-perceived accuracy of both the real-world outer door scene and the virtual world outer door scene for a limited period, and it gradually deviates with time.
- However, user test results show that the experience drone pilots agree with the simulated drone's physical realism and manoeuvrability. Moreover, there is a deviation between the user perceived position/speed of the simulated drone and the actual position/speed of the simulated drone.



- According to the evaluation results, early-stage predictions provided by the proposed and developed drone simulator are substantially accurate with rudder variations. Hence, drone piloting drills/missions/exercisers that require a short period (less than 25 seconds) for the entire activity can be simulated with substantial user perceive realism, behavioural realism and physical realism.

There is physical and behavioural realism in the proposed machine learning-based drone simulation environment. However, a wide range of further research work can be carried out to improve existing physical and behavioural realism. Some of the most critical and selected future research works are described below.

- Perform a wide range of drone manoeuvres and collect larger data sets that cover more drone dynamics and enhance behavioural realism by increasing the accuracy of the machine learning-based drone dynamic model.
- Design and develop a cylindrical or spherical display system, connect head-mounted display (HMD) and carry out experiments to enhance the physical realism and user perceive accuracy.

#### REFERENCES

- [1] D. Sandaruwan, M. Jayasundara, N. Kodikara, and S. Pitigala, "Machine learning based approach to simulate drone dynamics related to figure of eight maneuvering pattern," *European Journal of Computer Science and Information Technology*, vol. 7, no. 5, pp. 16–25, 2019.
- [2] K. Nonami, "Research and development of drone and roadmap to evolution," *Journal of Robotics and Mechatronics*, vol. 30, no. 3, pp. 322–336, 2018.
- [3] J. F. Keane and S. S. Carr, "A brief history of early unmanned aircraft," *Johns Hopkins APL Technical Digest*, vol. 32, no. 3, pp. 558–571, 2013.
- [4] J. D. Blom, "Unmanned aerial systems: A historical perspective," 2010, available online at: <https://www.armyupress.army.mil/Portals/7/combat-studies-institute/csi-books/OP37.pdf>.
- [5] A. Intwala and Y. Parikh, "A review on vertical take off and landing (vtol) vehicles," *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, vol. 2, no. 2, pp. 187–191, 2015.
- [6] Parrot Drone SAS, "Parrot ar.drone 2.0 elite edition," 2018, available online at: <https://www.parrot.com/global/drones/parrot-ar-drone-20-elite-edition>.
- [7] DJI Enterprise, "Drone solutions for a new generation of work," 2018, available online at: <https://www.dji.com/phantom-4>.
- [8] 3D Robotics Inc, "Scale your drone operations with enterprise atlas," 2018, available online at: <https://3dr.com>.
- [9] W. Hoffman, "Drones will need air traffic control, stanford and nasa will build it," *inverse innovation*, 2016, available online at: <https://www.inverse.com/article/13951-drones-will-need-air-traffic-control-stanford-and-nasa-will-build-it>.
- [10] T. Kelly, "The booming demand for commercial drone pilots," 2017, available online at: <https://www.theatlantic.com/technology/archive/2017/01/drone-pilot-school/515022/>.
- [11] Pilot Institute, "How much salary can a drone pilot expect?" 2020, available online at: <https://pilotinstitute.com/drone-pilot-salary/>.
- [12] Det Norske Veritas (DNV), "Standard for certification no. 2.14 maritime simulator systems," 2011.
- [13] D. Sandaruwan, N. Kodikara, C. Keppitiyagama, R. Rosa, M. Jayawardena, and P. Samarasinghe, "User perception of the physical & behavioural realism of a maritime virtual reality environment," in *2012 UKSim 14th international conference on computer modelling and simulation*. IEEE, 2012, pp. 172–178.
- [14] D. Sandaruwan, "Computer based ship model to predict real-time ship motion characteristics for perception enhanced virtual environment," 2015.
- [15] D. Sandaruwan, N. Kodikara, C. Keppitiyagama, R. Rosa, K. Dias, R. Senadheera, and K. Manamperi, "Low cost immersive vr solutions for serious gaming," in *Algorithmic and Architectural Gaming Design: Implementation and Development*. IGI Global, 2012, pp. 407–429.
- [16] P. Wang, Z. Man, Z. Cao, J. Zheng, and Y. Zhao, "Dynamics modelling and linear control of quadcopter," in *2016 International Conference on Advanced Mechatronic Systems (ICAMEchs)*. IEEE, 2016, pp. 498–503.
- [17] C. Caceres, D. Amaya, and J. M. Rosario, "Simulation, model and control of a quadcopter ar drone 2.0," *dynamics*, vol. 8, p. 10, 2016.
- [18] F. Sabatino, "Quadrotor control: modeling, nonlinear control design, and simulation," 2015.
- [19] C. MASSÉ, O. GOUGEON, D.-T. NGUYEN, and D. SAUSSIÉ, "Modeling and control of a quadcopter flying in a wind field: A comparison between lqr and structured h $\infty$  control techniques," in *2018 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2018, pp. 1408–1417.
- [20] Spark Pilots, "Dji assistant 2 ver 1.1.2," 2018, available online at: <https://sparkpilots.com/threads/dji-assistant-2-ver-1-1-2.11976/>.
- [21] DJI, "Dji download center," 2018, available online at: <https://www.dji.com/downloads/software/assistant-dji-2>.
- [22] Horizon Hobby, LLC, "Realfight 8 horizon hobby edition," 2018, available online at: <https://www.realfight.com>.
- [23] DroneSim Pro, "Dronesim pro drone simulator," 2019, available online at: <https://www.dronesimpro.com/>.
- [24] Immersion RC, "Liftoff – the drone race simulator," 2019, available online at: <https://www.immersionrc.com/fpv-products/liftoff-drone-race-simulator>.
- [25] Heli-X, "Professional r/c flight simulation," 2015, available online at: <https://www.heli-x.info/cms/>.
- [26] K. Amer, M. Samy, M. Shaker, and M. ElHelw, "Deep convolutional neural network based autonomous drone navigation," in *Thirteenth International Conference on Machine Vision*, vol. 11605. SPIE, 2021, pp. 16–24.
- [27] P. T. Jardine, S. Givigi, and S. Yousefi, "Parameter tuning for prediction-based quadcopter trajectory planning using learning automata," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 2341–2346, 2017.
- [28] N. Smolyanskiy, A. Kamenev, J. Smith, and S. Birchfield, "Toward low-flying autonomous mav trail navigation using deep neural networks for environmental awareness," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 4241–4247.
- [29] P.-H. Chen and C.-Y. Lee, "Uavnet: an efficient obstacle detection model for uav with autonomous flight," in *2018 International Conference on Intelligent Autonomous Systems (ICoIAS)*. IEEE, 2018, pp. 217–220.
- [30] A. Loquercio, A. I. Maqueda, C. R. Del-Blanco, and D. Scaramuzza, "Dronet: Learning to fly by driving," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 1088–1095, 2018.
- [31] R. P. Padhy, S. Verma, S. Ahmad, S. K. Choudhury, and P. K. Sa, "Deep neural network for autonomous uav navigation in indoor corridor environments," *Procedia computer science*, vol. 133, pp. 643–650, 2018.
- [32] F. Sadeghi and S. Levine, "Cad2rl: Real single-image flight without a single real image," *arXiv preprint arXiv:1611.04201*, 2016.
- [33] J. Hwangbo, I. Sa, R. Siegwart, and M. Hutter, "Control of a quadrotor with reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 2, no. 4, pp. 2096–2103, 2017.
- [34] O. Çakir and T. Yüksel, "Neural network control for quadrotors," *American Academic Scientific Research Journal for Engineering, Technology, and Sciences*, vol. 31, no. 1, p. 191–200, May 2017.
- [35] D. Jeong, M. Baek, and S.-S. Lee, "Long-term prediction of vehicle trajectory based on a deep neural network," in *2017 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2017, pp. 725–727.
- [36] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Icml*, 2010.

- [37] J. Brownlee, "A gentle introduction to the rectified linear unit (relu)," 2019, available online at: <https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/>.
- [38] R. D. Jackson, M. Jump, and P. L. Green, "Predicting on-axis rotorcraft dynamic responses using machine learning techniques," *Journal of the American Helicopter Society*, vol. 65, no. 3, pp. 1–12, 2020.
- [39] J. Requeima, W. Tebbutt, W. Bruinsma, and R. E. Turner, "The gaussian process autoregressive regression model (gpar)," in *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 2019, pp. 1860–1869.
- [40] A. Punjani and P. Abbeel, "Machine learning for helicopter dynamics models," Ph.D. dissertation, University of California, Berkeley, 2014.
- [41] Stanford University, "The stanford university autonomous helicopter," 2010, available online at: <http://heli.stanford.edu>.
- [42] J. Zhang, X. Yang, Z. Jin, and L. Li, "Distance estimation in virtual reality is affected by both the virtual and the real-world environments," *i-Perception*, vol. 12, no. 3, p. 20416695211023956, 2021.
- [43] C. J. Ziemer, J. M. Plumert, J. F. Cremer, and J. K. Kearney, "Estimating distance in real and virtual environments: Does order make a difference?" *Attention, Perception, & Psychophysics*, vol. 71, no. 5, pp. 1095–1106, 2009.
- [44] R. M. Baños, C. Botella, A. Garcia-Palacios, H. Villa, C. Perpiñá, and M. Alcaniz, "Presence and reality judgment in virtual environments: a unitary construct?" *CyberPsychology & Behavior*, vol. 3, no. 3, pp. 327–335, 2000.
- [45] DJI Technology company, "Phantom 4 specs," 2016, available online at: <https://www.dji.com/phantom-4/info>.
- [46] DJI Technology Company, "Phantom 4 pro specs," 2016, available online at: <https://www.dji.com/phantom-4-pro/info>.
- [47] J. Vrana and R. Singh, "Nde 4.0—a design thinking perspective," *Journal of nondestructive evaluation*, vol. 40, no. 1, pp. 1–24, 2021.
- [48] Priya Pedamkar, "Differences between machine learning vs neural network," available online at: <https://www.educba.com/machine-learning-vs-neural-network/>.
- [49] SAS Institute Inc, "Artificial neural networks: What they are & why they matter," 2022, available online at: [https://www.sas.com/en\\_us/insights/analytics/neural-networks.html](https://www.sas.com/en_us/insights/analytics/neural-networks.html).
- [50] R. Bhatia, "When not to use neural networks," 2018, available online at: <https://medium.datadriveninvestor.com/when-not-to-use-neural-networks-89fb50622429>.
- [51] M. Kuhn, K. Johnson *et al.*, *Applied predictive modeling*. Springer, 2013, vol. 26.
- [52] P. Sharma, "Improving neural networks – hyperparameter tuning, regularization, and more (deeplearning.ai course #2)," 2018, available online at: <https://www.analyticsvidhya.com/blog/2018/11/neural-networks-hyperparameter-tuning-regularization-deeplearning/>.
- [53] Tensorflow.org, "Tensorflow," 2015, available online at: <https://github.com/tensorflow/tensorflow>.
- [54] Keras, "Keras api reference," (n.d.), available online at: <https://keras.io/api>.
- [55] R. S. Ferguson, *The Principles of Quantum Mechanics*. CRC Press, 2014.
- [56] Epic Games, "The world's most open and advanced real-time 3d creation tool," (n.d.), available online at: <https://www.unrealengine.com/en-US/>.
- [57] OGRE, "Open source 3d graphics engine," (n.d.), available online at: <https://www.ogre3d.org>.
- [58] Unity, "Unity real-time development platform," (n.d.), available online at: <https://unity.com>.
- [59] VDrift, "Vdrift," (n.d.), available online at: <http://vdrift.net>.
- [60] A. Mairaj, A. I. Baba, and A. Y. Javaid, "Application specific drone simulators: Recent advances and challenges," *Simulation Modelling Practice and Theory*, vol. 94, pp. 100–117, 2019.
- [61] D. Sandaruwan, "Computer based ship model to predict real-time ship motion," Ph.D. dissertation, University of Colombo, Colombo, 2015.
- [62] Microsoft, "Airsim," 2017, available online at: <https://microsoft.github.io/AirSim/>.
- [63] Microsoft Research, "Aerial informatics and robotics platform," (n.d.), available online at: <https://www.microsoft.com/en-us/research/project/aerial-informatics-robotics-platform/>.
- [64] B. There, "Concepts, effects and measurements of user presence in synthetic environment," 2003.
- [65] T. Tsiatsos, K. Andreas, and A. Pomportsis, "Evaluation framework for collaborative educational virtual environments," *Journal of Educational Technology & Society*, vol. 13, no. 2, pp. 65–77, 2010.
- [66] S. McLeod, "Likert scale definition, examples and analysis," 2019, available online at: <https://www.simplypsychology.org/likert-scale.html>.

# Stacking Deep-Learning Model, Stories and Drawing Properties for Automatic Scene Generation

Samir Elloumi<sup>1</sup>, Nzamba Bignoumba<sup>2</sup>  
University of Jeddah, Jeddah, Saudi Arabia<sup>1</sup>  
Tallinn University of Technology, Estonia<sup>2</sup>

**Abstract**—Text-image mapping is of great interest to the scientific community, especially for educational purposes. It helps young learners, mainly those with learning difficulties, to better understand the content of stories. In this paper, we propose to capture the teacher’s experience in manually building relevant scenes for animal behavior stories. This manual work, which consists of a pair of texts and a set of elementary images, is fed into a Long Short-Term Memory (LSTM) followed by a Conditional Random Field (CRF) that aims to associate the relevant words in the text with their corresponding elementary image while preserving the drawing properties. This association is then used for scene construction. Several experiments were conducted to show how better the constructed scenes convey textual information than the scenes constructed from the competitor’s models.

**Keywords**—Text to image conversion; elementary image; image composition; deep-learning; drawing properties

## I. INTRODUCTION

The use of simple illustrative approaches by instructors to facilitate and make new concepts easier to comprehend dates back to when kids first start school. For example, they have used pins to stick the images on woody boards to explain actions, verbs or any other pedagogical purpose. A set of elementary images (EIs) such as a boy, a ball, a tree, to name but a few, were carefully preserved in the drawers and reused to compose new images or scenes as needed. Consequently, the main goal of this research is to replicate EI composition.

Text to image mapping has witnessed a great interest for the scientific community, in particular for the educational purposes and early learners, especially those who have reading disabilities. Many multimedia systems proposed to visually explain topics, news streams or stories by annotating news articles with pictures [1], enriching textual content [2], [3] or composing scenes [4]. Nevertheless, there are some common limitations in the existing systems which have not been properly addressed:

- Several existing multimedia systems can retrieve pictures automatically from the image search engine and generate illustrations [5], [6], [7], [8]. However, manual work is required to filter out inappropriate pictures, which reveals the excessive manual efforts behind these systems as indicated in [4].
- Multimedia systems for illustrating Arabic text are very limited, which reflects the current technical difficulties in understanding Arabic text.
- Many pictures are available on the web, but they lack textual descriptions or captions to be included in a relevant image search.

- Most of existing multimedia systems can illustrate text based on retrieved images. However, none of them, to our best knowledge, considers extracting elementary objects from retrieved images and using them for composing new pictures from scratch.

Although the use of a multimedia repository (MR) is one of the most common approaches to building scenes from sentences, obtaining the appropriate images from this repository to construct scenes is far from being an easy task. In some cases, we may not even be able to find accurate images that match the input sentence. However, the MR may contain EI images whose assembly may perfectly match the input sentence. The main objective of our work is then to propose a model that builds scenes with EIs while preserving the relevant implicit or explicit information contained in the sentence.

To achieve this, we implemented a model based on Recurrent Neural Networks (RNN) coupled with the Conditional Random Field (CRF) model to first identify in the input sentence words that correspond to EIs and secondly, predicting the dimensions and positions of these EIs in a such a way that the information contained in the input sentence is preserved. The EIs names and their respective dimensions and positions, which are represented in matrix form, will be transmitted to a system that will be responsible for building the scene. We implement our model and building system with python because it contains very advanced machine learning libraries such as Keras and TensorFlow. The sentences we are addressing are those relating to animal behavior.

The remainder of the paper is organized as follows. We first provide a literature review in Section II and then elaborate on our proposed method in Section III. We present experimental results in Section IV and evaluation in Section VI. Finally, we conclude this paper and discuss future directions in Section VII.

## II. RELATED WORK

Generating images from text (T2I), is an area of growing interest in computer science. Indeed, many approaches have been proposed, inspired by the way the human brain proceeds when trying to understand simple to more complex sentences. Based on his cognitive memory (human being), the comprehension of a simple text can be done by associating with each word of a text an image [9]. This process of understanding a text from an image has given rise to several approaches, which consist in generating scenes (set of images) from elementary annotated images. For instance, Coyne and Sproat [10], generate scenes from the WordsEye

image database by retrieving the images whose annotations match the words of the sentences. Rather than relying on all the words in the sentence to construct the scenes, Zhu et al. [11], proposed another approach that consists of identifying the most relevant concepts in the sentence and constructing the scene by merging the pictorial representations of these concepts. As the position of the elementary images may play a salient role in adequately conveying the meaning of the sentence, Yamada et al. [12], proposed a geometric model where the scenes are built considering the spatial constraints of the object described in the text. While these aforementioned approaches work well in practice for simple sentences, they quickly find their limit with complex ones. Indeed, their image database is not exhaustive, i.e. some words do not have their corresponding images. Moreover, even if one assumes to have an exhaustive image database, some abstract words like “lying, politics” cannot be represented by an image. To overcome this limitation, Rada et al. [13], implemented a system that generates scenes from complex sentences by coupling text and images to overcome this limitation. Although this last approach attempts to solve the abstract word problem, it is also limited by its exhaustive image database. It should be mentioned that all the approaches mentioned so far do not deal with spatial constraint and abstract word issues simultaneously. Moreover, the goal of T2I systems is to generate realistic scenes with exclusively images (which can be difficult when the embedded images have different backgrounds). In order to propose a model that addresses these issues simultaneously, the researchers turned to a deep learning model called GAN, proposed by Goodfellow et al. [14].

Originally proposed to generate realistic images by learning pixel distribution from a train image dataset, GAN is made up of two adversarial neural networks: a generator  $G$  and a discriminator  $D$ .  $G$  is trained to generate images by learning the distribution of real images and fooling the discriminator, in contrast  $D$  is trained to identify which images are generated or real. The spectacular results, obtained with GAN, have generated enormous enthusiasm in the creation of models derived from the latter. Thus, in order to improve the MNIST digit generation, Mirza et al. [15], proposed a conditional GAN (cGAN) where the generator and discriminator are conditioned by a class label  $y$ . Inspired by this approach for the T2I task, rather than conditioning the generation process by a  $y$ -class label, Reed et al. [16], proposed to condition it by the whole sentence embedding obtained from a pre-trained text encoder. Compared to [16] where the generated images had a resolution of  $64 \times 64$ , the authors in [17], proposed a Tac-gan-text conditioned auxiliary classifier generative adversarial network (TAC-GAN) capable of generating a higher resolution image i.e.  $128 \times 128$ . In order to improve image resolution, another paradigm based on generators and discriminators stacking has emerged. Zhang and al. in [18], proposed a model called StackGAN composed of two generative stages. The first stage is dedicated to the generation of a coarse  $64 \times 64$  pixel image given a random noise vector and textual conditioning vector, while the second produces an image of  $256 \times 256$ . An improved version of [18], composed of three stacked generators and discriminators, was proposed in [19]. To avoid stacking several pairs of discriminators and generators layers, other approaches like [20] [21] proposed to reduce the number of generator and increase the number of discriminator (or vice versa). As

with these stacked listed generative models, a generated image is dependent on a previous one (except for the initial), the poor quality of the latter can lead to an inaccurate generated image. Therefore, to prevent this from happening, the authors proposed Dynamic Memory Generating Adversarial Networks (DM-GANs) in which a dynamic memory unit is designed to select important textual information based on the content of the initial generated image and then use it to generate the next image. In addition to the features extracted from the generated image, the authors in [22], integrated also aspect-level features (processed from the input text) to update and enhance word-level feature in order to refine the next image.

By making the assumption that an image generated from text should be based on the relevant words in addition to the whole sentence (attention), Xu et al. in [23], built a fine-grained text to image generation with Attentional Generative Adversarial Networks (AttnGAN). Huang et al. [24], proposed a grid-based attention model that involves applying an attention mechanism between object-grid regions and word phrases. Similar to our work, part-of-speech tagging is applied to extract word features. Other models applying attention in different fashion like [25], [26], [27], [28] were also proposed.

Initially proposed to solve signature and face verification problems, Siamese networks designed with two branches (split-parameter neural networks) processing a pair of inputs have been also repurposed for the T2I task in [29], [26]. In both, each branch takes as input a text (caption) and generates an image. However, in [29] the objective loss function is employed to minimize / maximize the distance between the features extracted in each branch to learn a semantically meaningful representation, depending on whether the two captions are from the same ground truth image (intra-class pair) or not (inter-class pair) in [26], the objective loss function aims to minimize the feature distance between generated image and corresponding ground truth image while maximizing the distance to another real image associated with a different caption. In [30], the authors proposed a model derived from Siamese networks called Text-SeGAN in which negative sampling of image pairs is carried out with several strategies so that the model is able to detect the most subtle differences between two images and therefore improve the generation process. Another T2I approach, called cycle-consistent image generation by re-description architectures inspired by CycleGAN [31], was also implemented in [32], [33]. The principle of this approach is to learn a semantically consistent representation between text and image by appending a captioning network and train the network to produce a semantically similar caption from the generated image.

Unlike the CycleGAN derived models where the image generation process is conditioned by some inputs, unconditional generative models [34], [35], [36] were built upon unconditional image generation models [37], [38] for T2I purpose. For example, in [35], the authors proposed a model called textStyleGAN in which the text is previously passed through a pre-trained image-to-text matching network to compute the embedded representation of the whole text as well as the words of this last. These embedded representations are combined with noise and fed into textStyleGAN to generate an image. Instead of only generating images from texts, other approaches have added additional supervision tricks. For example, to

generate complex images (images with several objects also called scenes) from Microsoft Common Objects in Context (MS COCO), Sharma et al. proposed ChatPainter [39] where, in addition to scene captions, they also rely on dialogues (pairs of question-answers) describing the scenes. Other works, in [40], [41], [42], were also relying on dialogue approach. The authors, in [43], [44], used multiple captions to iteratively improve the image quality.

Although GAN models are tending to democratize to the detriment of classical approaches (due to their much higher performance requiring less and less human intervention), the generation of complex images remains a challenge. In order to overcome this challenge, we proposed a model based on a combination of the classical approach, probabilistic graphical model and deep learning models. Each of these approaches contributes as follows:

- The classic approach consists of using annotated images from a Multimedia Repository (MR).
- Deep learning is to explore the sequential pattern while encoding each word in the text.
- And probabilistic graphical model to match the latent word representation emitted by the deep learning model to their corresponding part-of-speech tag;

The novelty of our proposal which, is the introduction of new types of part-of-speech tags that describe the action performed by a specific object and define the coordinates of the latter in the scene. This part-of-speech tagging step of associating each word of the text with a tag based on the action, position and dimension of the object in the scene is the basis of our system.

### III. OUR APPROACH: SCENE BUILDER BASED ON ELEMENTARY IMAGES

As depicted in Fig. 1, there are three phases in our builder system. In phase 1, we need to prepare the dataset as an input for the tag matching phase. In phase 2, the model performing the tag matching stage, is fed with the word-tokens of the sentences (stories) and their corresponding tags (object names, positions and sizes) processed in phase 1. Finally, in phase 3, a scene builder is implemented to build the scenes from the input sentences, based on the tag matrices obtained in the tag matching stage.

#### A. Phase 1: Learning Dataset Preparation

We have developed a graphical tool that allows an instructor to write a story's text and manually draws its related scene by inserting Elementary images (EIs) in a graphical drawing area. Based on his expertise, the instructor selects the most relevant images and arranges them in the graphical area to fit the story's meaning. We consider that this expertise is the key point on which we based our approach. Hence, we collect all instructor actions, namely the selected images, their positions and sizes, and we save them in a database. While choosing an image from the toolbox, the user selects its related textual parts in the story. This is considered as an implicit image annotation that serves later to map a text into an image. In the following, we give more details about the tools and the EIs repository already prepared in [45].

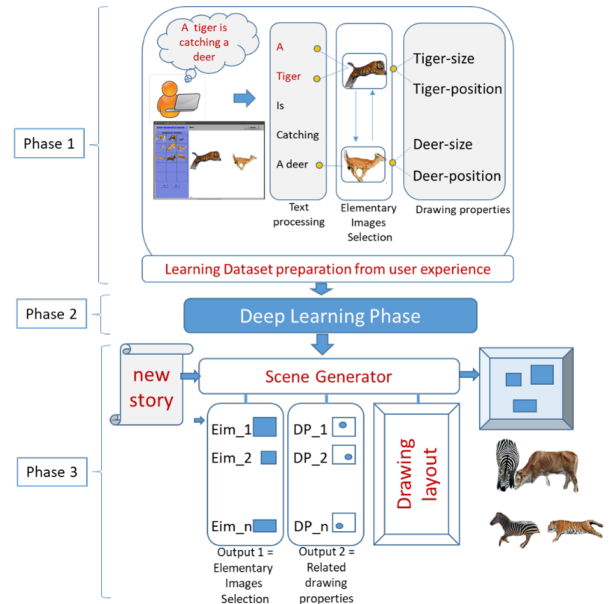


Fig. 1. System architecture.

1) *Elementary Images Repository*: An initial EIs repository was constructed in [45]. It contains around 1540 EIs collected from Google image and ImageNet. The EIs were obtained by following a particular process. In fact, based on existing online libraries such as Google Images, ImageNet, etc., a set of images were collected and stored in a local folder. For each image, the model Mask R-CNN [46] object extraction tool was applied in order to obtain EIs with some of their drawing properties. Subsequently, an image captioning process was applied on extracted elementary objects in order to automatically assign a caption for each one of them.

2) *The Tool*: In our Image Story Generator tool, image composing is designed in two ways: manual image composing and automatic image composing. In our current tool version, we compose new images or pictures manually, allowing thereby flexible working with the tool. We briefly describe how a user composes new images using our tool. First, a user input keywords in an input field on the top of the tool main interface and hits enter. Note, we use single keyword only at this current version. The retrieved EIs from EMR are displayed in a panel on the Graphical Toolbox on the left side of the main interface, as indicated in Fig. 2. The user or the teacher can drag and drop the main interface, locate images and resize them, thereby composing a new scene describing the input sentence. The teacher can successively search for other EIs doing same steps as described. Thus, the teacher can show the final illustration to the students. The newly created image is stored locally. Therefore, the system extracts the image sizes, positions, etc. This information is saved as drawing properties and will be further used for generating new pictures/images dynamically.

3) *Input preparation for the tag matching stage*: In this phase, we arrange the database content to fit the deep learning model dedicated to the tag matching stage.

Once getting a manually satisfactory scene, i.e., arranging the EIs and setting their sizes, positions as it should be,

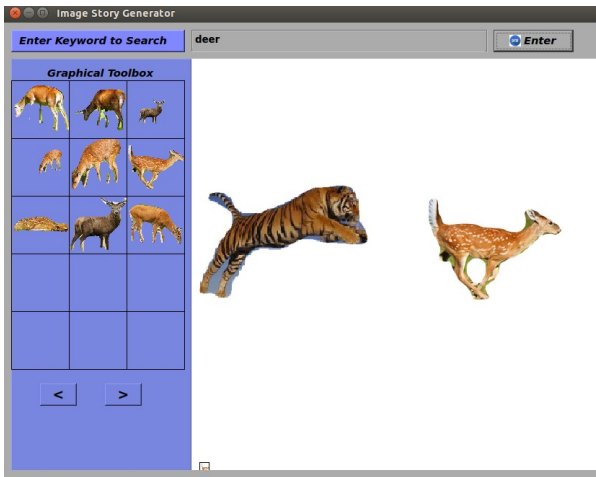


Fig. 2. Scene drawing tool

the prepared image is saved. More specifically, the metadata related to this composite image such as the names of EIs, their respective position and dimension, within the global one image, their links to some parts of the initial text are saved as well.

The saved file will act as our data set. It will then be processed by a deep learning model so that from a sentence  $\mathbf{t} = [t_0 \ t_1 \ \dots \ t_{n-1}]$  where  $t_n$  is the token code for the  $n^{th}$  word in  $\mathbf{t}$ , a corresponding matrix tag  $\mathbf{M}_{n,m}$  is generated.  $n$  is the number of words in the sentence  $\mathbf{t}$  and  $m$  is the number of tag types. Hence, the matrix is defined by eq. 1:

$$\mathbf{M} = \begin{matrix} t_0 \\ \vdots \\ t_{n-1} \end{matrix} \begin{bmatrix} \hat{y}_{0,0} & \hat{y}_{0,1} & \cdots & \hat{y}_{0,m-1} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{y}_{n-1,0} & \hat{y}_{n-1,1} & \cdots & \hat{y}_{n-1,m-1} \end{bmatrix} \quad (1)$$

$y_{n,m}$  is the tag  $m$  associated with the sentence's word at position  $n$ . For our study,  $m$  is equal to 3 since we attempt to generate the word type (whether it is an object or not), it's size, and it's position. Note that, we used the Beginning, Inside, Outside (B-I-O) notation, for the tag's representation with reference to the "Named Entity Recognition" (NER) information extraction technique [47]. This technique allows capturing the action performed by an object (if animal type) in the sentence. For each word in a sentence, we associate a three tags vector. The first value of this tag vector is an image tag that determines whether the image is an object and identifies which word describes the action performed by the object. The second one is the position tag, which determines the position in which the *EI* will be placed. The third one, which is the size tag, determines the size that the *EI* will have within the scene. The Table I illustrates an example of a phrase and its different tags. It is worth mentioning that the Inside notation is only applied to image tag generation because we want to identify the action performed by an animal-type *EI*. Which is not a necessity for position and size tags, which are spatial tags.

- The values of the "Image Tag" column allows identifying the names of the EIs in the sentence as well as the actions they perform (if any). The names and actions

TABLE I. A SENTENCE WITH THREE TAGS

Sentence	Image Tag	Position Tag	Size Tag
A	'O'	'O'	'O'
lion	'B-object'	'B-w1Lh2B'	'B-large'
lying	'I-object'	'O'	'O'
in	'O'	'O'	'O'
the	'O'	'O'	'O'
savannah	'B-object'	'B-w0Rh0B'	'B-entire'
observes	'O'	'O'	'O'
a	'O'	'O'	'O'
gazelle	'B-object'	'B-w1Rh1B'	'B-middle'
.	'O'	'O'	'O'

of these objects will make it possible to construct the path to retrieve the appropriate *EI*. If we take for example the sentence of Table I, the paths obtained will be:  $\dots/lying/lion$ ;  $\dots/savannah$ ;  $\dots/gazelle$ . So the *EIs* that will constitute the scene will be a lying lion, a gazelle and the savannah. The savannah object like other objects such as the street, the river etc. present in our sentences are background objects. They are associated with the size tag "B-entire".

- The "Position Tag" column allows defining the position of *EIs* in the scene. Rather than considering the exact positions in pixel, we have considered regions landmark. In fact, the entire image is discretized or split in  $n \times p$  parts (where  $n$  and  $p$  are positive numbers). Suppose that the image dimension is  $w \times h$  (where  $w$  represents its width and  $h$  its height respectively in pixels). The scene (i.e. the entire image) is discretized in  $\frac{w}{n} = w_1 + w_2 + w_3 + \dots + w_n$  width parts and  $\frac{h}{p} = h_1 + h_2 + h_3 + \dots + h_p$  length parts. Based on that, each *EI* is associated with a position tag of type  $B - w_x H h_y V$ . Hence,  $w_x$  indicates that the image should be positioned from the  $x^{ieth}$  width part of the scene, starting in a  $H$  direction where  $H \in \{L, R\}$  (L for left, R for right) and  $h_y$  to indicate to the system that the image should be positioned from the  $y^{ith}$  part in the height of the scene, starting in a direction  $V$  where  $V \in \{T, B\}$  (T for top, B for bottom). In the example shown in Fig. 3, the image is divided in  $n = 6$  width parts and  $p = 5$  height parts. In this scene we have three *EI* which are: the lion, the gazelle and the savannah in the background. To each of these images, the following position tags are respectively assigned to them:  $B - w_1 L h_2 B$ ,  $B - w_1 R h_1 B$ ,  $B - w_0 L h_0 B$ .
- The values of the "Size Tag" column allows defining the size of the EIs within the scene. In this column, we have 4 different tags. Each one of them is associated with a predefined width and height.

The next step is the learning model preparation.

### B. Phase 2: Tag Matching

The tag matching stage involves associating a triplet of tags with each word of the sentence. Due to their ability of data extraction and high accuracy in classification and

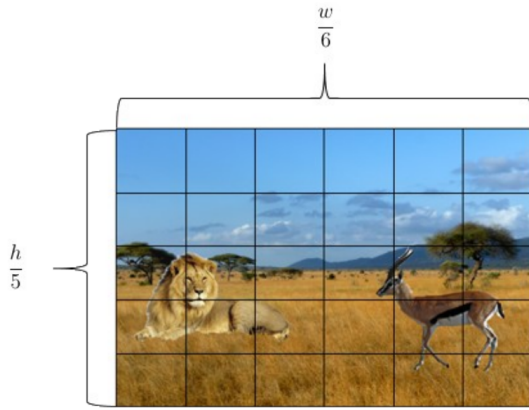


Fig. 3. Discretized image

prediction tasks [48], [49], we used stacked deep learning models (Embedding, Bi-LSTM and Dense layers) to perform the data extraction in the tag matching stage. Technically, we subsequently pass our sentence input through an Embedding layers, then into a Bi-LSTM layer and finally into a Dense layer wrapped by a TimeDistributed layer. Their respective roles are as follows:

- Embedding layer: it aims to associate with each word of the sentence, a vector of real numbers encoding the semantics of the latter and its relationship with the other words.
- Bi-LSTM: it re-encodes vectors outputted from the Embedding layer to another vectors of real numbers for which the encoding process leverage on the sequential pattern existing in the input text.
- Dense layer: It extracts the information contained in the vectors obtained from the Bi-LSTM layer. It is wrapped by a TimeDistributed because a simple Dense layer can only be fed by a single vector. In our case, we have as many vectors as there are words in the sentence, that is why we used a TimeDistributed layer.

As CRF [50] has achieved leading results in speech part tagging, we appended it to our model for the tag matching stage. Technically, the output of the Dense layer is passed through three different CRF layers where, each one is dedicated to matching image, position, and size tags. The whole architecture of the tag matching stage is depicted in Fig. 4.

These steps will be repeated during the training phase in order to define the optimal parameters of the model. The loss function for parameter optimization is defined as follows:

$$loss = loss_{image} + loss_{position} + loss_{size} \quad (2)$$

$$loss_{tag} = -(\ln(p(\mathbf{y}|\mathbf{t})_f) + \ln(p(\mathbf{y}|\mathbf{t})_b)) \quad (3)$$

$$p(\mathbf{y}|\mathbf{t}) = \frac{e^{S(\mathbf{y},\mathbf{t})}}{\sum_{\mathbf{y}' \in \mathbf{Y}} e^{S(\mathbf{y}',\mathbf{t})}} \quad (4)$$

$$S(\mathbf{y}, \mathbf{t}) = \sum_{i=0}^{N-1} \mathbf{A}_{y_i, y_{i+1}} + \sum_{i=1}^{N-1} \mathbf{P}_{i, y_i} \quad (5)$$

Where  $tag \in \{image, position, size\}$ ,  $\mathbf{t}$  the input sentence vector, and  $\mathbf{y}$  the corresponding tag vector.  $f$  stands for forward-pass and  $b$  for backward-pass in the Bi-LSTM layer.  $S$  is the cross-score function between words in  $\mathbf{t}$  and tags in  $\mathbf{y}$ .  $\mathbf{A}$  is a matrix where coefficients are the transition score from  $y_i$  to  $y_{i+1}$ .  $\mathbf{P}$  is a matrix whose coefficients are the scores of the pairs  $(t_i, y_i)$ . Once the matrix is generated, a mapping between its values and all the  $EIS$  to build our scenes is performed. In the next subsection, we formally describe how scenes are constructed.

### C. Phase 3: Scene Construction

In phase 3, the scene constructor system is ready to map a new story to a scene by mapping the words of the input sentence which their corresponding images base on the tags in  $\mathbf{M}$ . It is formally described as follows:

$$f(\mathbf{t}, \mathbf{M}, EIS) = \hat{S} \quad (6)$$

Where  $\hat{S}$  represents the scene built from the sentence  $\mathbf{t}$ . The first objective of the function  $f$ , is to map each word (if they are objects), to their corresponding elementary image  $EI \in EIS$ . Second, each  $EI$  will be positioned in the scene respecting its generated size and position. The Pseudo-Algorithm 1 presents the mains steps for Scene Constructor.

---

#### Algorithm 1 Scene Constructor $f$

---

**Require:**  $\mathbf{M}, \mathbf{t}, EIS$

**Ensure:**  $\hat{S}$

- 1:  $\hat{S} \leftarrow \text{init}(\text{Matrix})$   $\triangleright$  We initialize the scene, i.e. we create an image with only white pixels.
  - 2:  $\mathbf{Tags} \leftarrow \text{getTags}(\mathbf{M})$   $\triangleright$   $\mathbf{Tags}$ , is a matrix where each row  $i$  represents the path ( $\mathbf{Tags}[i, 0]$ ), the position ( $\mathbf{Tags}[i, 1]$ ) and size ( $\mathbf{Tags}[i, 2]$ ) of an object in  $\mathbf{t}$ . The path is processed based on the generated image tags.
  - 3: **for**  $i, \mathbf{tgs} \leftarrow \text{enumerate}(\mathbf{Tags})$  **do**
  - 4:      $EI \leftarrow \text{getEI}(\mathbf{tgs}[0], EIS)$
  - 5:     **if**  $\mathbf{tgs}[1] \neq \text{"B-entire"}$  **then**
  - 6:          $\text{Draw}(\hat{S}, EI, \mathbf{tgs}[1], \mathbf{tgs}[2])$   $\triangleright$  This function draws EI in the scene according to their position and their size.
  - 7:     **else**
  - 8:          $k \leftarrow i$
  - 9:     **end if**
  - 10: **end for**
  - 11:  $EI \leftarrow \text{getEI}(\mathbf{Tags}[k, 0], EIS)$   $\triangleright$  After saving the index  $i$  associated with the object with B-integer as size tag, we draw this object at the end to avoid overlapping other EIs.
  - 12:  $\text{Draw}(\hat{S}, EI, \mathbf{Tags}[k, 1], \mathbf{Tags}[k, 2])$
  - 13: **return**  $\hat{S}$
- 

## IV. EXPERIMENTAL STUDY

In this section, we first present the dataset used to evaluate our model and the hyperparameters retained after a grid search. Second, we benchmark the model against state-of-the-art competitors. For the implementation, we used Python machine learning libraries: TensorFlow and Keras. NumPy,

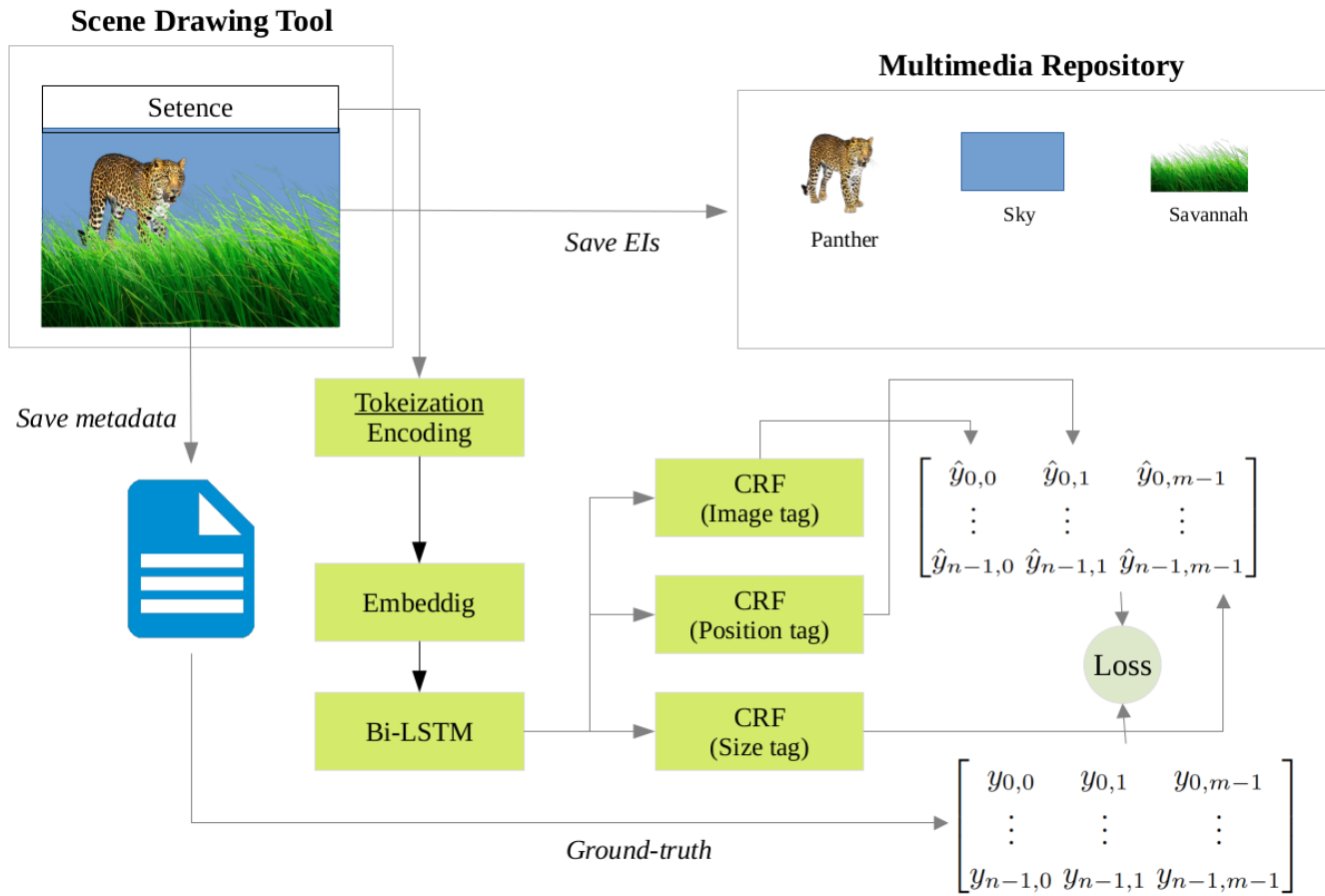


Fig. 4. Tags prediction model architecture: learning phase.

Pandas and Matplotlib were also used for the preprocessing stage and the scene construction stage.

#### A. Dataset

The dataset we used (see Table II) consists of 210 sentences that describe the behavior of 52 animals and their interactions with 34 objects that are not animals (e.g. vegetables, wood, etc.). We used 190 sentences for training (90%) and 21 sentences for testing (10%). The sentences are the same ones we used to manually build scenes and record metadata in the graphical tool. They are organized in such a way that the same action can be associated with different EIs of the animal type. This allows the model to better understand that actions are more related to words representing EIs of animal type. Take the example of the sentences “Lions eat meat” and “Cats eat meat in front of the door”. These two sentences are similar. Lions and cats, which are EIs of animal type, perform the same action: eating. Thus, the model will figure out that the action of eating is more related to animal-type EIs than to other EIs in the sentence.

#### B. Hyperparameters

After a grid search, the different hyperparameters of the layers that make up our model were set as follows:

##### a) Embedding layer:

- The number of distinct words in the corpus 166.
- The dimension of the vector space into which each word initially encoded with the one-hot encoding technique will be projected is set to 40.
- The fixed length of the input sentence is 100.
- The Boolean parameter that specifies whether the token 0 is a padding token or not is set to *True*. This is because the input sentences are of different lengths.

##### b) BI-LSTM layer:

- The number of cells in each of our LSTM forward and LSTM backward networks is 100.
- The Boolean parameter allowing to indicate whether the BI-LSTM network should return either a sequence of token (in vector forms) or just one token is set to *True* because, we need to match a tag for each word in the input sentence.
- The recurrent dropout, which is a regularization technique that prevents over-fitting [51], [52] is set to 0.7. This means that the probability that a Bi-LSTM cell is skipped during training is 0.7.



TABLE II. TRAINING SET DESCRIPTION

	Total number of objects	Total number of distinct objects
Animal objects	258	52
Other objects	153	34
Total	411	86

c) *Time Distributed:*

- the TimeDistributed layer is a layer that wraps each Bi-LSTM output in a Dense layer. The number of cells in this Dense layer is set to 100. We used the *Relu* activation function in this layer.

d) *CRF:*

- The CRF parameter we set here is the number of tags that could match for each EI in the input sentence. Each EI can have 3 image tags which corresponds to the number of units in the CRF layer dedicated to the image tag; 54 position tags which corresponds to the number of units in the CRF layer dedicated to the position tag; 5 size tag which corresponds to the number of units in the CRF layer dedicated to the size tag.

For the training phase, we set the number of epochs to 300 and the batch size to 3. We used the gradient descent optimizer [53], [54], to update the model weights.

In the next section, we present the results obtained, and the metrics used to evaluate them.

### C. Experimental Results

To evaluate the tag generation component, we considered the F1-score metric [55] computed from precision and recall. They are defined as follows :

- Precision: Designates the number of classes different from the class “O” which are correctly predicted by the system divided by the total number of positive classes predicted by the system [55].

$$\frac{|true\ positives|}{|true\ positives| + |false\ positives|} \quad (7)$$

- Recall: Designates the number of classes different from the class “O” which are correctly predicted by the system out of the total number of classes which are not “O” [55].

$$\frac{|true\ positives|}{|true\ positives| + |false\ negatives|} \quad (8)$$

- The F1-score: Designates the ratio of the product of the recall by the precision on their sum [55].

$$F_{\beta} = (1 + \beta^2) \frac{PR}{\beta^2 P + R} \quad (9)$$

Where  $\beta = 1$  determines the balance coefficient between precision and recall. The results, in Table III, summarize the F1 score obtained for the image, position and size generation.

We can see that they are quite satisfactory. The lower F1 score obtained with the position tag generation component is explained by its high number of tags, which makes it less stable. On the other hand, due to their low number of tags, the image and size tag components have higher F1 scores. We hypothesize that the F1 score of the image tag component is better than that of the size tag component (which has a similar number of tags) due to the use of the Inside notation, which reinforces the generation process.

The evaluation we performed with the F1 score just allowed us to find the optimal hyperparameters for the generation of the tag matrices. In the next section, we present the scenes built from these tag matrices.

## V. BUILDING SCENES

To build scenes, we will couple our tag prediction model to our scene building tool. The latter will take as input a sentence  $t_i$ , its corresponding tag matrix (the one predicted by the model)  $M_i$  as well as all of our EIs. For application, 21 test sentences and their corresponding tag matrix are considered. The results obtained are presented in Tables IV, V.

Among the 21 sentences, 18 images correctly reflect the corresponding input texts. So we have a success rate of 18/21 or 85.7%. However, if our system succeeds in: placing EI in positions that reflect the interactions between them; producing very realistic scenes thanks to a position label class that allows the entire background to be painted with an EI and; clearly representing the action performed by the EI, errors and anomalies are to be underlined. First, we can see in Table V line 1, that the predicted tag matrix contains errors. The model classifies the word “next” as an action performed by the object “elephant”, which is inconsistent. Secondly, in Table V line 2, we see that for the dimension tag, the class assigned to “fly” is “B-entire” so the latter is considered as a background EI. Therefore, since its dimension is smaller than the background, it cannot cover it completely. In the opposite case (if it has the same dimension as the background), it would make no sense to have a “fly” as a background image. Although errors are observed when predicting the tags, the model is still able to provide images that are more or less correlated with the sentences.

After evaluating our model with our corpus, we will evaluate it with another corpus and compare the results obtained against those of state-of-the-art text-to-image systems and models.

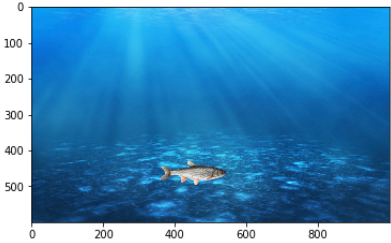


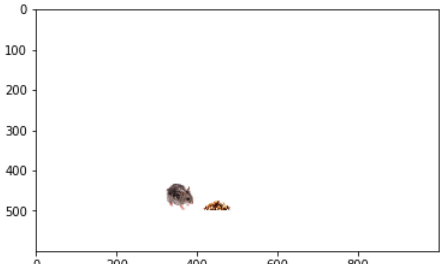
## VI. EVALUATION OF THE MODEL AGAINST THE GOOGLE IMAGE SEARCH ENGINE, THE MULTIMEDIA SYSTEM AND A GAN MODEL

This approach, which consists of building scenes manually rather than automatically from elementary images, has already

TABLE III. F1 SCORE OF IMAGE, POSITION AND TAG GENERATION COMPONENT

Metric	Image tag	Position tag	Size tag	Average
F1-score(%)	78.2	60.2	70.9	<b>69.77</b>

TABLE IV. SENTENCES, THEIR TAG MATRIX AND CORRESPONDING SCENES

Sentences $t_i$	Tags Matrix $M_i$	Scene
$\begin{bmatrix} A \\ \text{fish} \\ \text{in} \\ \text{the} \\ \text{sea} \\ . \end{bmatrix}$	$\begin{bmatrix} O & O & O \\ B - \text{object} & B - \text{small} & B - w3Lh2B \\ O & O & O \\ O & O & O \\ B - \text{object} & B - \text{entire} & B - w0Lh0B \\ O & O & O \end{bmatrix}$	
$\begin{bmatrix} \text{Horses} \\ \text{in} \\ \text{the} \\ \text{fence} \\ . \end{bmatrix}$	$\begin{bmatrix} B - \text{object} & B - \text{middle} & B - w2Lh3B \\ O & O & O \\ O & O & O \\ B - \text{object} & B - \text{entire} & B - w0Lh0B \\ O & O & O \end{bmatrix}$	
$\begin{bmatrix} \text{Bears} \\ \text{in} \\ \text{the} \\ \text{forest} \\ . \end{bmatrix}$	$\begin{bmatrix} B - \text{object} & B - \text{middle} & B - w3Lh1B \\ O & O & O \\ O & O & O \\ B - \text{object} & B - \text{entire} & B - w0Lh0B \\ O & O & O \end{bmatrix}$	
$\begin{bmatrix} A \\ \text{mouse} \\ \text{eat} \\ \text{nuts} \\ . \end{bmatrix}$	$\begin{bmatrix} O & O & O \\ B - \text{object} & B - \text{small} & B - w2Lh2B \\ I - \text{object} & O & O \\ B - \text{object} & B - \text{small} & B - w3Lh2B \\ O & O & O \end{bmatrix}$	

been proposed. [45]. In order to evaluate our model, we will compare the results obtained by this multimedia system to see

whether automatic construction can be as efficient as manual construction. In addition, we will also compare it to Google

TABLE V. SENTENCES, THEIR TAG MATRIX AND RELATED SCENES, PART 2

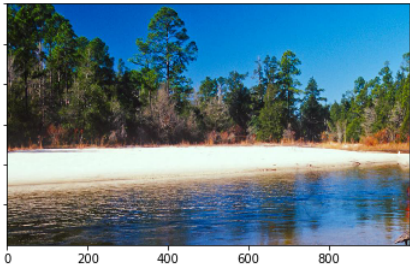
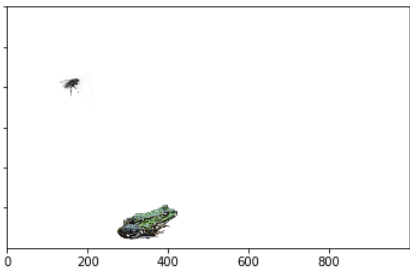
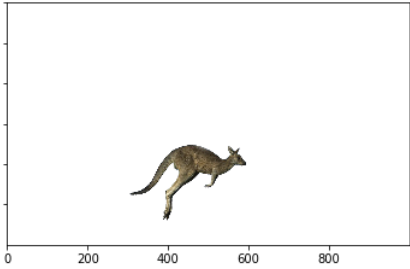

Sentence $t_i$	Tag Matrix $M_i$	Scene
$\begin{bmatrix} A \\ \text{elephant} \\ \text{lying} \\ \text{next} \\ \text{to} \\ \text{a} \\ \text{river} \\ . \end{bmatrix}$	$\begin{bmatrix} O & O & O \\ B - \text{object} & B - \text{large} & B - w2Lh2B \\ I - \text{object} & O & O \\ I - \text{object} & O & O \\ O & O & O \\ O & O & O \\ B - \text{object} & B - \text{entire} & B - w0Lh0B \\ O & O & O \end{bmatrix}$	
$\begin{bmatrix} A \\ \text{frog} \\ \text{observes} \\ \text{a} \\ \text{fly} \\ . \end{bmatrix}$	$\begin{bmatrix} O & O & O \\ B - \text{object} & B - \text{middle} & B - w2Lh1B \\ O & O & O \\ O & O & O \\ B - \text{object} & B - \text{entire} & B - w1Rh2B \\ O & O & O \end{bmatrix}$	
$\begin{bmatrix} A \\ \text{kangaroo} \\ \text{jumping} \\ . \end{bmatrix}$	$\begin{bmatrix} O & O & O \\ B - \text{object} & B - \text{middle} & B - w3Lh2B \\ I - \text{object} & O & O \\ O & O & O \end{bmatrix}$	
$\begin{bmatrix} A \\ \text{bear} \\ \text{plunged} \\ \text{in} \\ \text{the} \\ \text{river} \\ . \end{bmatrix}$	$\begin{bmatrix} O & O & O \\ B - \text{object} & B - \text{large} & B - w3Lh2B \\ I - \text{object} & O & O \\ O & O & O \\ O & O & O \\ B - \text{object} & B - \text{entire} & B - w0Lh0B \\ O & O & O \end{bmatrix}$	

image and to the GAN model.

#### A. Corpus of Sentences

To evaluate our model, the chosen corpus of sentences will be used to evaluate the multimedia system [45]. This choice is

simply justified by the fact that our model is an improvement of the previous one. Instead of building the scenes manually, as is the case in this multimedia system, we will build them automatically. The number of sentences contained in this corpus is 20. All relating to animal behavior. We will take

8 for our evaluation. It should nevertheless be noted before the evaluation that our model does not yet take into account the cardinality of EIs. For example, in the following sentences, “two birds” and “three flowers”, the model cannot draw 2 birds or 3 flowers.

### *B. Results*

In this section, we present the different results provided by the multimedia system, Google image, the GAN model and our system. As far as the multimedia system and Google image are concerned, we will simply retrieve the results from [45]. For each sentence we associate its related image as provided by each system. The results are depicted in Tables VI, VII, VIII, IX and X.

For the comparison analysis, we consider three criteria. The first one is Principal actors (PA). It allows evaluating to what extent the objects present in the sentence are present in the generated image. The second criterion, the event (EV), consists in evaluating to what extent the event described in the text is represented in the generated image. The third and last criterion assess to what extent the spatial positions of the objects and the temporal aspect (S & T) are respected. Each of these evaluations will be out of five as indicated in the Table XI. Thereafter, we will calculate an average specific to each sentence for each model and for each model we will calculate the average of each criterion obtained for all eight sentences.

We can see through the Table XI that our model outperforms both the Google Image system and the GAN model. The latter is the worst one. However, still our previous multimedia system [45] gives the best results since the construction of the scenes is done manually. However, still the multimedia system [45] gives the best results since the construction of the scenes is done manually. However, our model can be used for short texts that might be addressed to a large young learners students.

TABLE VI. RESULTS FOR EACH SYSTEM (PART 1)

Phrases	Google image	Multimedia System	GAN Model	Our System
The tiger is running behind a herd of deer			No Object Found in the System	
The two hungry bears were wandering through the forest				
The crocodile grapples with zebras				
Zebras, bulls and tiger drink water from the lake				

TABLE VII. RESULTS FOR EACH SYSTEM PART 2

Phrases	Google image	Multimedia System	GAN Model	Our System
The camels walked and crouched on the sand				
Elephants smashed the rabbit's house				
The horses and the zebras run behind the elephants in the stadium				
A group of cats jumps behind small butterflies				

TABLE VIII. RESULTS FOR EACH SYSTEM PART 3

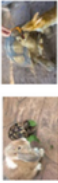


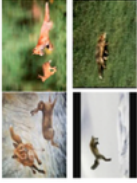




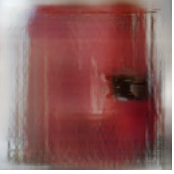

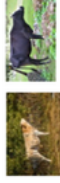



Phrases	Google image	Multimedia System	GAN Model	Our System
The rabbits and the turtles eat lettuce.			No Object Found in the System	
The fox is running behind the little rabbit.			No Object Found in the System	
The horses and the bulls are running in the playground.				
The bulls eat leaves of trees.				

TABLE IX. RESULTS FOR EACH SYSTEM PART 4

Phrases	Google image	Multimedia System	GAN Model	Our System
Two tigers and one lion eat meat.			No Object Found in the System	
The dogs and the foxes run behind the deer.				 
The turtles dive in the lake.			No Object Found in the System	
The two bears stand on two trees with behives.	 			



TABLE X. RESULTS FOR EACH SYSTEM PART 5





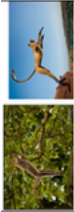






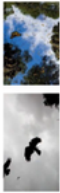



Phrases	Google image	Multimedia System	GAN Model	Our System
The crows are standing on a high tree.				
The monkey jumped over sleeping rabbits.			No Object Found in the System	
The camels eat grass.				
The crows flew with the butterflies over the tree.				

TABLE XI. RESULTS W.R.T THE FOURTH CRITERIA MAIN ACTORS (AP), MAIN EVENT (EP), SPATIAL AND TEMPORAL COMPOSITION (SP) AND SCENE GLOBAL COMPOSITION

Id	Google image			Multimedia System			GAN Model			Our System						
	AP	EV	S & P	Global	AP	EV	S & P	Global	AP	EV	S & P	Global				
1	4.5	3.2	2.2	3.3	5	3.7	3.2	4	0	0	0	0	5	4	4	4.3
2	4	4	4	4	5	5	5	5	2	2	2	2	3	4	4	3.7
3	4.5	4.5	4.5	4	4	4	3	3.6	1	0.5	0	0.5	5	4	4	4.3
4	2.2	3.7	4	3.3	5	5	5	5	1	1	1	1	5	5	5	5
5	3.7	2.5	4.2	3.5	5	5	5	5	1	0	0	0.33	5	4	4	4
6	5	1	0	2	4.5	2.2	2.2	3	2	1	0	1	4	0	3.5	2.5
7	4	4	2	3.3	5	5	5	5	2	0	1	1	5	3.5	5	4.5
8	3	4	4	3.6	5	4	5	4.7	0	0	0	0	4	4	4	4
9	3.5	2.2	2.7	2.8	5	3	4	4	0	0	0	0	4	2	3.5	3.2
10	5	5	5	5	5	5	5	5	0	0	0	0	4	4	4	4
11	3.2	3.7	4	3.7	5	5	5	5	0	0	0	0	4	2.5	4	3
12	3	3.7	3.7	3.5	5	5	5	5	0	0	0	0	3.5	3	4	3.5
13	5	5	5	5	5	5	5	5	0	0	0	0	3	4	5	4
14	3	2	2	2.3	5	5	5	5	0	0	0	0	3	2.5	3	2.8
15	5	5	5	5	5	5	5	5	0	0	0	0	5	5	5	5
16	5	5	5	5	5	5	5	5	1	0	1	0.7	4	2	3	3
17	4.2	4.7	3.2	4.1	5	5	4.5	4.8	0	0	0	0	5	4	5	4.3
18	3	5	2.5	3.5	4	4	4	4	0	0	0	0	3.5	5	5	4.5
19	5	4.5	5	4.8	5	5	5	5	0.5	0	0	0.16	5	3	5	4.3
20	5	5	4.5	4.8	4	4	4	4	0	0	0	0	3	1	2	2
<b>Average</b>	<b>4</b>	<b>4.1</b>	<b>3.6</b>	<b>3.8</b>	<b>4.8</b>	<b>4.5</b>	<b>4.5</b>	<b>4.6</b>	<b>0.5</b>	<b>0.2</b>	<b>0.2</b>	<b>0.3</b>	<b>4.2</b>	<b>3.3</b>	<b>4.1</b>	<b>3.8</b>

## VII. CONCLUSION

In this paper, we have discussed how we could capture a user expertise to produce a new tool for automatic Stories illustration. The requirement is an available set of elementary images representing different types of objects that a user (or a tutor) uses from a toolbox and draw a scene related to a given story. The originality of our work is that we suggest capturing the “know-how” of the tutor and make it transfer to a deep-learning based model. The latter is coupled with a drawing tool in order to build automatically stories illustration by images compositions. We have conducted several experiments to find the optimal values for the deep-learning model such as: “Loss function”, “Batch size”, “Number of epochs”, etc. In addition, comparisons with existing approaches that make scene generation were also presented. Obviously, the quality of the “manually” constructed scenes remains the best one, nevertheless, our new approach gives very interesting results.

Even though we obtain satisfactory results, we agree that our approach face some limitations. One of them is that the Mask-CNN that we used to extract the EIs is not trained in end-to-end fashion. In our future work, we plan to train it in this way so that the model becomes more accurate. As we noticed in the results Subsection VI-B, in some cases the model fails to construct the scene (empty background) as the appropriate EI was not available on our Multimedia Repository. Therefore, we also plan to diversify the sources of image databases so that the MASK-CNN can extract more EIs. Due to the complexity of the sentences, we specifically focused our work on sentences related to animal’s behaviors. We then plan to incrementally integrate sentences related to other facts to make the model more generic.

## ACKNOWLEDGMENT

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-22-DR-96). The authors, therefore, acknowledge with thanks the University of Jeddah for its technical and financial support.

## REFERENCES

- [1] F. M.-N. A. Ramisa, F. Yan and K. Mikolajczyk, “Breakingnews:article annotation by image and text processing,” *arXiv preprint arXiv:1603.01354*, 2016.
- [2] R. Agrawal, S. Gollapudi, A. Kannan, and K. Kenthapadi, “Enriching textbooks with images,” in *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 2011, pp. 1847–1856.
- [3] A. Vatani, M. Taleby, and M. Rahimi, “An effective automatic image annotation model via attention model and data equilibrium,” *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 3, 2018.
- [4] R. Agrawal, S. Gollapudi, A. Kannan, and K. Kenthapadi, “Vishit: A visualizer for hindi text,” *Proceedings - 2014 4th International Conference on Communication Systems and Network Technologies, Bhopal*, pp. 886–890, 2014.
- [5] S. Aramini, E. Ardizzone, and G. Mazzola, “Automatic illustration of short texts via web images,” in *Proceedings of the 6th International Conference on Information Visualization Theory and Applications (IVAPP-2015)*, 10-14 January 2015.
- [6] D. Delgado, J. M. a. e. s., and N. Correia, “Automated illustration of news stories,” in *Proceedings of the 2010 IEEE Fourth International Conference on Semantic Computing, Pittsburgh*, vol. II, 2010, pp. 1035–1040.
- [7] A. B. Goldberg, J. Rosin, X. Zhu, and C. R. Dyer, “Toward text-to-picture synthesis,” in *NIPS 2009 Symposium on Assistive Machine Learning for People with Disabilities*, 2009.
- [8] H. Li, J. Tang, G. Li, and T.-S. Chua, “Word2image: Towards visual interpretation of words,” in *MM’08 - Proceedings of the 2008 ACM International Conference on Multimedia, with co-located Symposium and Workshops, Vancouver*, 2008, pp. 813–816.
- [9] M. C. Potter, J. F. Kroll, B. Yachzel, E. Carpenter, and J. Sherman, “Pictures in sentences: understanding without words,” *Journal of Experimental Psychology: General*, vol. 115, no. 3, p. 281, 1986.
- [10] C. Bob and R. Sproat, “Wordseye: an automatic text-to-scene conversion system,” in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 2001, pp. 487–496.
- [11] X. Zhu, A. B. Goldberg, M. Eldawy, C. R. Dyer, and B. Strock, “A text-to-picture synthesis system for augmenting communication,” in *AAAI*, vol. 7, 2007, pp. 1590–1595.
- [12] A. Yamada, T. Yamamoto, H. Ikeda, T. Nishida, and S. Doshita, “Reconstructing spatial image from natural language texts,” in *COLING 1992 Volume 4: The 14th International Conference on Computational Linguistics*, 1992.
- [13] R. Mihalcea and C. W. Leong, “Toward communicating simple sentences using pictorial representations,” *Machine translation*, vol. 22, no. 3, pp. 153–173, 2008.
- [14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [15] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” *arXiv preprint arXiv:1411.1784*, 2014.
- [16] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, “Generative adversarial text to image synthesis,” *arXiv preprint arXiv:1605.05396*, 2016.
- [17] A. Dash, J. C. B. Gamboa, S. Ahmed, M. Liwicki, and M. Z. Afzal, “Tac-gan-text conditioned auxiliary classifier generative adversarial network,” *arXiv preprint arXiv:1703.06412*, 2017.
- [18] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, “Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 5907–5915.
- [19] —, “Stackgan++: Realistic image synthesis with stacked generative adversarial networks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1947–1962, 2018.
- [20] L. Gao, D. Chen, J. Song, X. Xu, D. Zhang, and H. T. Shen, “Perceptual pyramid adversarial networks for text-to-image synthesis,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, 2019, pp. 8312–8319.
- [21] Z. Zhang, Y. Xie, and L. Yang, “Photographic text-to-image synthesis with a hierarchically-nested adversarial network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6199–6208.
- [22] S. Ruan, Y. Zhang, K. Zhang, Y. Fan, F. Tang, Q. Liu, and E. Chen, “Dae-gan: Dynamic aspect-aware gan for text-to-image synthesis,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 13 960–13 969.
- [23] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He, “Attngan: Fine-grained text to image generation with attentional generative adversarial networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1316–1324.
- [24] W. Huang, R. Y. Da Xu, and I. Oppermann, “Realistic image generation using region-phrase attention,” in *Asian Conference on Machine Learning*. PMLR, 2019, pp. 284–299.
- [25] B. Li, X. Qi, T. Lukasiewicz, and P. Torr, “Controllable text-to-image generation,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [26] H. Tan, X. Liu, X. Li, Y. Zhang, and B. Yin, “Semantics-enhanced adversarial nets for text-to-image synthesis,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 10 501–10 510.

- [27] H. E. M. Shamardan, "All in focus image generation based on new focusing measure operators," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 12, 2016. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2016.071217>
- [28] S. M. and R. Aarathi, "Text to image gans with roberta and fine-grained attention networks," *International Journal of Advanced Computer Science and Applications*, vol. 12, 01 2021.
- [29] G. Yin, B. Liu, L. Sheng, N. Yu, X. Wang, and J. Shao, "Semantics disentangling for text-to-image generation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2327–2336.
- [30] M. Cha, Y. L. Gwon, and H. Kung, "Adversarial learning of semantic relevance in text to image synthesis," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, 2019, pp. 3272–3279.
- [31] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
- [32] Q. Lao, M. Havaei, A. Pesaranhader, F. Dutil, L. D. Jorio, and T. Fevens, "Dual adversarial inference for text-to-image synthesis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7567–7576.
- [33] T. Qiao, J. Zhang, D. Xu, and D. Tao, "Mirrorgan: Learning text-to-image generation by redescription," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1505–1514.
- [34] D. M. Souza, J. Wehrmann, and D. D. Ruiz, "Efficient neural architecture for text-to-image synthesis," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.
- [35] D. Stap, M. Bleeker, S. Ibrahim, and M. ter Hoeve, "Conditional image generation and manipulation for user-specified content," *arXiv preprint arXiv:2005.04909*, 2020.
- [36] M. Yuan and Y. Peng, "Bridge-gan: Interpretable representation learning for text-to-image synthesis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 4258–4268, 2019.
- [37] A. Brock, J. Donahue, and K. Simonyan, "Large scale gan training for high fidelity natural image synthesis," *arXiv preprint arXiv:1809.11096*, 2018.
- [38] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4401–4410.
- [39] S. Sharma, D. Suhubdy, V. Michalski, S. E. Kahou, and Y. Bengio, "Chatpainter: Improving text to image generation using dialogue," *arXiv preprint arXiv:1802.08216*, 2018.
- [40] T. Niu, F. Feng, L. Li, and X. Wang, "Image synthesis from locally related texts," in *Proceedings of the 2020 International Conference on Multimedia Retrieval*, 2020, pp. 145–153.
- [41] S. Frolov, S. Jolly, J. Hees, and A. Dengel, "Leveraging visual question answering to improve text-to-image synthesis," in *Proceedings of the Second Workshop on Beyond Vision and Language: inTEgrating Real-world kNowledge (LANTERN)*, 2020, pp. 17–22.
- [42] S. A. J. Zaidi, A. Buriro, M. Riaz, A. Mahboob, and M. Riaz, "Implementation and comparison of text-based image retrieval schemes," *International Journal of Advanced Computer Science and Applications*, 2019.
- [43] K. Joseph, A. Pal, S. Rajanala, and V. N. Balasubramanian, "C4synth: Cross-caption cycle-consistent text-to-image synthesis," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 358–366.
- [44] J. Cheng, F. Wu, Y. Tian, L. Wang, and D. Tao, "Rifegan: Rich feature generation for text-to-image synthesis from prior knowledge," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10911–10920.
- [45] S. Elloumi, J. M. AlJa'am, and J. Zakraoui, "Building multimedia repository for composing images perspective," *SN Applied Sciences*, vol. 1, no. 9, p. 1116, 2019.
- [46] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
- [47] V. Yadav and S. Bethard, "A survey on recent advances in named entity recognition from deep learning models," *arXiv preprint arXiv:1910.11470*, 2019.
- [48] Y. Tang and A. Duan, "Using deep learning to predict the east asian summer monsoon," *Environmental Research Letters*, vol. 16, no. 12, p. 124006, 2021.
- [49] W. Zhang, S. Yan, J. Li, X. Tian, and T. Yoshida, "Credit risk prediction of smes in supply chain finance by fusing demographic and behavioral data," *Transportation Research Part E: Logistics and Transportation Review*, vol. 158, p. 102611, 2022.
- [50] J. Laerty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of ICML*, 2001.
- [51] T. Poggio, K. Kawaguchi, Q. Liao, B. Miranda, L. Rosasco, X. Boix, J. Hidary, and H. Mhaskar, "Theory of deep learning iii: explaining the non-overfitting puzzle," *arXiv preprint arXiv:1801.00173*, 2017.
- [52] X. Sun, X. Ren, S. Ma, and H. Wang, "meprop: Sparsified back propagation for accelerated deep learning with reduced overfitting," in *Proceedings of the 34th International Conference on Machine Learning—Volume 70*. JMLR. org, 2017, pp. 3299–3308.
- [53] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.
- [54] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng, "On optimization methods for deep learning," in *Proceedings of the 28th International Conference on International Conference on Machine Learning*. Omnipress, 2011, pp. 265–272.
- [55] L. Derczynski, "Complementarity, f-score, and nlp evaluation," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016, pp. 261–266.

# A Machine Learning Hybrid Approach for Diagnosing Plants Bacterial and Fungal Diseases

Ahmed BaniMustafa<sup>1\*</sup>, Hazem Qattous<sup>2</sup>, Ihab Ghabeish<sup>3</sup> and Muwaffaq Karajeh<sup>4</sup>

Data Science & Artificial Intelligence Department, Isra University, Amman, Jordan<sup>1\*</sup>

Software Engineering Department, Princes Sumaya University for Technology, Amman, Jordan<sup>2</sup>

Department of Agricultural Sciences-As-Shoubak University College, Al-Balqa' Applied University, Al-Salt, Jordan<sup>3</sup>

Department of Plant Protection and Integrated Pest Management, Mutah University, Karak, Jordan<sup>4</sup>

**Abstract**—Bacterial and Fungal diseases may affect the yield of stone fruit and cause damage to the Chlorophyll synthesis process, which is crucial for tree growth and fruiting. However, due to their similar visual shot-hole symptoms, novice agriculturalists and ordinary farmers usually cannot identify and differentiate these two diseases. This work investigates and evaluates the use of machine learning for diagnosing these two diseases. It aims at paving the way toward creating a generic deep learning-based model that can be embedded in a mobile phone application or in a web service to provide a fast, reliable, and cheap diagnosis for plant diseases which help reduce the excessive, unnecessary, or improper use of pesticides, which can harm public health and the environment. The dataset consists of hundreds of samples collected from stone fruit farms in the north of Jordan under normal field conditions. The image features were extracted using a CNN algorithm that was pre-trained with millions of images, and the diseases were identified using three machine learning classification algorithms: (1) K-nearest neighbour (KNN); (2) Stochastic Gradient Descent (SGD); and (3) Random Forests (RF). The resulting models were evaluated using 10-fold cross-validation, with CNN-KNN achieving the best AUC performance with a score of 98.5%. On the other hand, the CNN-SGD model performed best in Classification Accuracy (CA) with a score of 93.7%. The results shown in the Confusion Matrix, ROC, Lift, and Calibration curves also confirmed the validity and robustness of the constructed models.

**Keywords**—Deep Learning; Machine Learning; Classification; Plant Diseases; Disease Diagnosis

## I. INTRODUCTION

In Jordan and other Mediterranean countries, agriculture participates with a high share of the annual GDP, and stone fruit trees are among the most important sources of crop yield [1]. However, Bacterial and Fungal diseases, which appear as shot holes in the leaves, may affect the yield of the trees and may eventually cause damage to the Chlorophyll synthesis process, which is crucial for plant growth and fruiting.

Shot hole disease is caused by several pathogens. The fungal shot hole, or the *Coryneum* blight, is caused by *Wilsonomyces carpophilus* (Lev.) [2]. This fungus is common in the Mediterranean region, but it also spreads in America, Australia, Africa, Asia, Oceania, and Europe [3]. It infects stone fruit trees of Prunus, mainly apricot, peach, nectarine, and cherry. However, the first three are the most commonly affected hosts [4], [5]. It damages twigs, buds, blossoms, fruits, and leaves. The damage is most noticeable on the leaves, however. The causal agent thrives in cool and moist conditions between summer and fall, mainly in early spring and anytime during wet

weather conditions [3], [6]. It develops rapidly under warmer temperatures [3] and overwinters on blossom buds, and cankers of branches [7]. The symptoms appear on the infected leaves as small, round reddish to purplish-brown specks with light green or yellow rings around them. Lesions can be circular to slightly ellipsoid; their tissues can become raised and scurfy and will tear along the lesion margins and may hang on at one attached point. It often dries up and falls away, giving the shot hole; as if someone fired a shotgun at the leaf [5].

As the disease spreads, more leaves get damaged until they fall. Significant infections can reduce photosynthesis, weaken the plant, and decrease the fruit yield eventually, which makes it a major concern for the stone fruit industry worldwide [8]. The other causal agent of the shot hole disease is the bacterium *Xanthomonas Arboricola* PV. *Pruni* (Xap) [9]. The symptoms of this bacterium are circular to irregular, water-soaked lesions on leaves. Later, lesions turn purple or brown. Usually, halos and cracks can be seen between the affected tissue and the surrounding healthy tissue. In the later stage, the infected tissue will be broken away under various natural forces, especially wind, and finally drop out, leaving a hole. Leaves with many holes or lesions will turn yellow or prematurely drop off [9]. It infects leaves, twigs, and fruits of the same stone fruit hosts of the *W. Carpophilus* fungus. The favourable conditions for the disease occurrence and spread are largely the same as the fungal shot hole, but it overwinters mainly in twig lesions (OEPP/EPPO, 2003). The disease is considered the most limiting factor to cherry Laurel production in landscapes and nursery production. Losses exceed 75% in the nursery due to this disease (unpublished data, J. Williams-Woodward). However, these diseases, which affect different parts of the trees and produce holes in leaves, are caused by different pathogens such as *Wilsonomyces Carpophilus* fungus and *Xanthomonas Arboricola* PV. *Pruni* bacteria. Sometimes it is difficult to differentiate the symptoms to identify the pathogen, even further, to identify if the symptoms are caused by a pathogen or insect feeding. This makes it difficult to specify suitable pesticides or implement a control plan to avoid damage to the trees.

Nonetheless, the infections caused by these diseases are difficult to identify and cure as they are usually confused with viral pest infections and with other soil and nutrition deficiencies. The accurate identification and diagnosis of these diseases infections through visual signs and symptoms may require deep knowledge and thorough experience, which is usually lacking in most ordinary farmers, and yet they may also

require laboratory tests and identification procedures, which are usually expensive to perform for most farmers. Yet, fighting these bacterial and fungal infections without a proper diagnosis may fail or may require using a wide spectrum of pesticides, placing additional financial costs which add to the short-term losses in the fruit yield and the long-term damages made to the trees, which may affect their lifespan. In addition, the use of pesticides may harm the trees and may also place concerns over food safety, which may cause several marketing challenges due to allowable limits imposed by food and drug administrations and may ultimately affect customer confidence. This all adds to other environmental concerns that are related to pesticide leaks into the soil and water and their short-term and long-term influence on wildlife.

Novice agriculturalists and farmers not knowing much about the disease may confuse the infections with other viral and pest diseases and deficiencies caused by factors related to soil, water, or other environmental conditions. They might also confuse these diseases with symptoms associated with other issues related to the incorrect use of pesticides and fertilisers. Therefore, the false identification and diagnosis of fungal and bacterial may waste farmers' time and effort in solving irrelevant issues. This in fact, lay additional financial pressure on farmers and may also cause other serious environmental and health-related issues, such as those related to the improper use of pesticides or unintentional insect feeding. Moreover, the spread of fungal and bacterial diseases in various environments may also contribute to the evolution of new races of causal agents with different symptoms, thus making them hard to diagnose under normal field conditions [6].

Machine learning techniques in general [10], and deep learning algorithms in particular such as convolutional neural networks [11], [12], [13] have recently witnessed impressive success in several scientific and commercial applications, particularly in the fields of computer vision, image recognition [14], [15], [16], and other classification applications [17], [18], [19]. One of the potential applications of machine learning is identifying and diagnosing plant diseases' infections [20], and their other related issues [21]. In this research, we have conducted an end-to-end empirical study that involves collecting, diagnosing, and identifying many stone fruit leaf samples that suffer from bacterial and fungal infections. The sample leaves have been photographed, categorised and then used to train and test three machine learning classifier models, which are created by combining convolutional neural networks with three classical machine learning algorithms: (1) K-nearest neighbour (KNN); (2) Stochastic Gradient Descent (SGD); and (3) Random Forests (RF). These algorithms were selected due to their reported success in image classification. All the constructed models will be evaluated using Classification Accuracy (CA), Area Under the Curve (AUC), Precision and F1 metrics, in addition, the Confusion Matrix, ROC, Lift and Calibration curves will also be used to confirm the validity and robustness of the created models.

The following four subsections are dedicated to investigating the related work and defining the research hypothesis, question, and aims. Section II provides details regarding the research methodology, materials and data collection, and an overview of the techniques applied and their performance metrics. Section III presents the results of model building and

evaluation which is followed by Section IV, which analyses the results obtained in this work in light of the research hypothesis, question, and objectives. It also compares the results with those reported by others. Section V provides a conclusion of the findings and comments of the result on the limitation of the study and highlights the contribution of this study. It also comments on the significance of the results and their potential applicability, as well as on the possible future work that is related to the study.

#### A. Related Work

In this subsection we investigate nine of the most popular related works and most cited studies that have been published in the last six years.

A study that was published in [22] reported 78% accuracy in classifying diseases that infect apple trees using the field-collected dataset, while [23] reported accuracy of 81% in classifying four grapes diseases using images that have been captured under field normal condition. On the other hand, [24] reported accuracy of 82% in classifying cucumber diseases using a research centner dataset, while [25] reported an accuracy of 83% in classifying disease that infects potato using images of their tubers which have been collected from potato fields. [26] reported yet another field-based study that involved classifying diseases that infect tea plants with an accuracy of 90%. [27] reported accuracy of 88% in classifying banana diseases using the plant village public dataset which is very close to the results reported by [28] which aimed at detecting and classifying diseases that infect tomato leaves and which also depended on using a public dataset that was downloaded from plant village. [29] reported an accuracy of 99% in classifying soybeans disease also using plant village public datasets. [30] also reported a 99% of classification accuracy using the plant village dataset, but this time for classifying twenty-six diseases that hit fourteen different crops. A recent study reported in [31] achieved classification accuracy of 96.63% using the ResNet-50 deep learning algorithm to predict 15 disease classes using 20,000 public dataset that is published by plant village. Another study published by [32] reported an accuracy of 96% using unsupervised learning. However, the study provides no details regarding the number of the samples in the dataset. Another study published in [33], reported an accuracy of 92.57% using YOLOv5 model for classifying 61 categories of plant diseases that hit ten different plant species. The dataset consisted of 36,258 images. A more recent study that used Multiple Linear Regression (MLP) reported 91% accuracy in identifying blister blight in tea plants[34], while another recent study reported the use of CNN algorithm to detect 13 different diseases achieved precision of 96.3% [35].

The analysis of the related work shows that most of the reported applied techniques scored a classification accuracy which ranged between 78% and 99%. It was noticed that the majority of the investigated works reported a classification accuracy above 90% depended on huge public datasets which involved thousands of images that have been captured in a controlled environment with a standardised camera type, images orientation, images resolution, images aspect ratio, lightning and sometimes the growth conditions of the plant itself. On the other hand, the studies that reported performance

of less than 90% depended on using a less number of images (a few hundred images) captured using ordinary cameras and for leaves of plants that were grown in field normal conditions. Furthermore, only few of the reviewed studies spent effort in diagnosing the plant diseases and investigating their cause and pathogens. Most of the other studies depended only on the disease's visual symptoms, which shade doubt on the validity of their findings. In addition, few of the surveyed studies commented on the utilisation and the practical use of their obtained results.

### B. Research Hypothesis

The research suggests that deep learning pre-trained algorithms can be used for the automatic identification of image features and then be classified using other machine learning classification algorithms to provide an accurate diagnosis of stone fruit fungal and bacterial diseases.

### C. Research Question

The research question seeks to answer the question: Can we build a successful machine learning model to diagnose stone fruit fungal and bacterial diseases?

### D. Research Aim

Creating a machine learning model for diagnosing stone fruit diseases will pave the way towards creating a generic deep learning-based model that can be embedded in a mobile phone application or a web service for the purpose of providing a fast, reliable, and cheap diagnosis of plant diseases, which helps to reduce the excessive, unnecessary, or improper use of pesticides in agriculture, which can be harmful to public health and the environment.

## II. MATERIALS AND METHODS

This section provides a brief description of the technology applied and the methodology implemented in order to automate and tackle the problem of stone fruit disease diagnosis in this research to tackle.

### A. Materials

The dataset consists of 500 sample images. 294 of the images belong to the Fungal class, while 106 images belong to the Bacterial class. 99 belong to the control healthy class. The dataset creation involved collecting and photographing hundreds of images of stone fruit that are infected with the bacterium *Xanthomonas arboricola* P.v. *Pruni* (*Xap*) and *Wilsonomyces Carpophilus* fungus.

The leaves were collected from various farms in the north of Jordan during the mid of August, which are located in a region with a moderate Mediterranean mountainous climate with an elevation between 630-950 Metres above sea level. The average temperature between April and September is 29.7 °C during the day and 17.2 °C, while the average humidity is 45% and the average pressure is 1008 mBar. The rain is very scarce during the data collection period. Fig 1 shows sample leaves that suffer from bacterial and fungal infections. The samples were collected and photographed in the field normal conditions.

The dataset has an equal number of images for each infection class. The stone fruit leaves were first collected and then photographed using a mobile phone camera with an 8M pixels resolution. The leaves were then examined by two domain experts who work as plant diseases experts and professors at al-Balqa applied University and Mutah University. The disease experts diagnosed the infection in each leaf and placed the leaves' images in their corresponding folders. The diagnosis of each leaf was then confirmed using plant disease literature and also by using the plant disease database sponsored by the European and Mediterranean Plant Protection Organisation (EPPO).

### B. Methods

The methods applied in this research involve performing three major steps: (1) The extraction of significant image features using deep learning convolutional neural networks VGG16 algorithms ; (2) the classification of images into Bacterial and Fungal classes using the KNN, SGD and RF machine learning algorithms; and (3) models evaluation based on 10-folds cross-validation and using confusion matrix, ROC, AUC, classification Accuracy and Precision metrics. Fig. 2 illustrates the three major steps in the proposed research methodology.

1) *Features Extraction*: The features in each image are extracted using the convolutional neural network VGG16 implementation that was pre-trained using 14 million images from the ImageNet database which correspond to 1000 labelled classes[36]. The algorithm was originally proposed by Simonyan and Zisserman to introduce several improvements to AlexNet which was developed earlier by google [37], [38].

The applied implementation was based on the Keras framework using Python and Caffe framework. Keras is a deep learning Python Library that focuses on simplifying deep learning model construction and visualising its using Tensor Flow, while Caffe is a deep learning framework that was developed by the BAIR AI research group at the University of California, Berkeley [39]. It enables constructing expressive, modularised and fast deep learning models [40], [41]

The constructed vectors for each image feature were then preprocessed, first by normalising the image features values within the interval between -1 and 1 and then, by removing the sparse features for each feature with missing values exceeding the ratio of 5%.

Deep learning is one of the most popular contemporary technology in machine learning that has recently witnessed unprecedented success in several applications, particularly in image and voice recognition. as it requires no or minimal feature engineering[11]. Deep learning is simply, a back propagation neural network algorithm that is based on creating a machine learning model using several layers, where each represents a level of abstraction[11].

Convolutional neural networks (CNN) are a class of deep learning algorithms that uses feed-forward neural networks to extract and learn image features using multiple layers[11], [42]. CNN has achieved success in several machine learning applications, particularly in image processing, recognition and classification. In convolutional neural networks, and each layer,

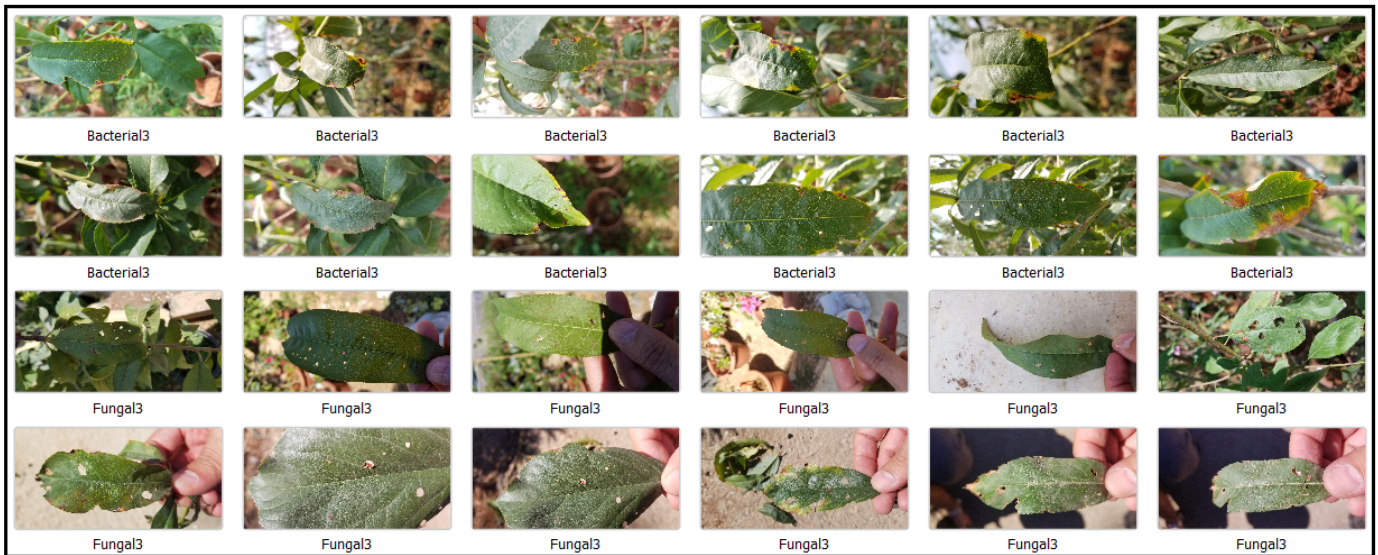


Fig. 1. Image samples that were collected and photographed in the field normal conditions for stone fruit leaves that suffer from fungal and bacterial infections.

the image dimensionality is reduced and its features are automatically identified, extracted, stored in vectors and then passed to a more specialised deeper layer which extracts even more features using its own "convolutions". A convolution does a job in terms of feature extraction that is analogous to light filters in optics[42], [43], [11]. Convolutional Neural Network (CNN) algorithms were used for extracting and analysing image features. The identified feature was then used for creating three machine learning classification models for identifying the stone fruit fungal and bacterial diseases using a dataset that consists of hundreds of images that have been captured for the infected leaves. Convolutions are used to learn the image features in a fashion analogous to filters, in each layer the dimensionality of the image is reduced using convolution and then fed into the lower layer [43]. The convolution at the top layer is used to filter, extract and learn the most generic features in the image and then pass them down toward the convolutions in the deeper layers that are used to filter, extract and learn the more specific features as illustrated in Fig. 2.

2) *Classification*: Machine learning (ML) concerns enabling computers to learn and improve from previous examples and prior experience using a wide spectrum of artificial intelligence techniques and algorithms[44], [45], [46]. Modern technology and recent advances in machine learning are the driving force behind what is called "The fourth industrial revolution."

Machine learning tries to mimic the ability of human beings and other intelligent species to learn and attain knowledge from prior experience and examples and then generalise and use this knowledge to respond to future situations and process unknown data. In machine learning classifiers, a model is built and then trained to predict the outcome or responses in a precooked dataset which is called "training data." The model is then validated by testing its ability to predict the unknown outcomes and responses in unknown datasets or situations, which is called "validation data."

Selecting the appropriate machine learning modelling technique is important and must take into consideration the nature of the dataset, study objectives, and the potential of the applied technique [46], [45], [47]. In this research, three algorithms have been applied to the targeted images and then evaluated to find the best machine learning algorithm that can be coupled with the deep learning convolutional neural networks VGG16 algorithm to diagnose the targeted leaves and classify their infection. The applied algorithms involve KNN, SGD, and RF. These algorithms will be introduced and discussed in detail in this section.

*K-Nearest Neighbours Algorithm (KNN)*: An instance-based learning technique that is used for classifying data samples by measuring their proximity to neighbouring data points that belong to a set of pre-labelled classes. This technique was first introduced by Evelyn Fix and JL Hodges, Jr. in their unpublished technical report while working at the USAF School of aviation [48]. KNN measures the Euclidean distance between predicted and training values belonging to a predefined class in a two-dimensional space using equation 1. The predicted classes for a point are determined based on a popularity vote regarding its distance from other neighbouring data points that belong to the neighbouring classes. KNN was used in several image classification applications that included classifying brain CT Scan [49] and recognising images of various objects [50]. In both applications, it achieved a classification accuracy of 80%.

$$\Delta(x_i, x_j) = \sqrt{\sum_{i=1}^n (|x_{in} - x_{jn}|)^2} \quad (1)$$

*Stochastic Gradient Descent (SGD)*: This algorithm classifies samples iteratively. In each iteration, the weights of the classification model are updated using equation 2.



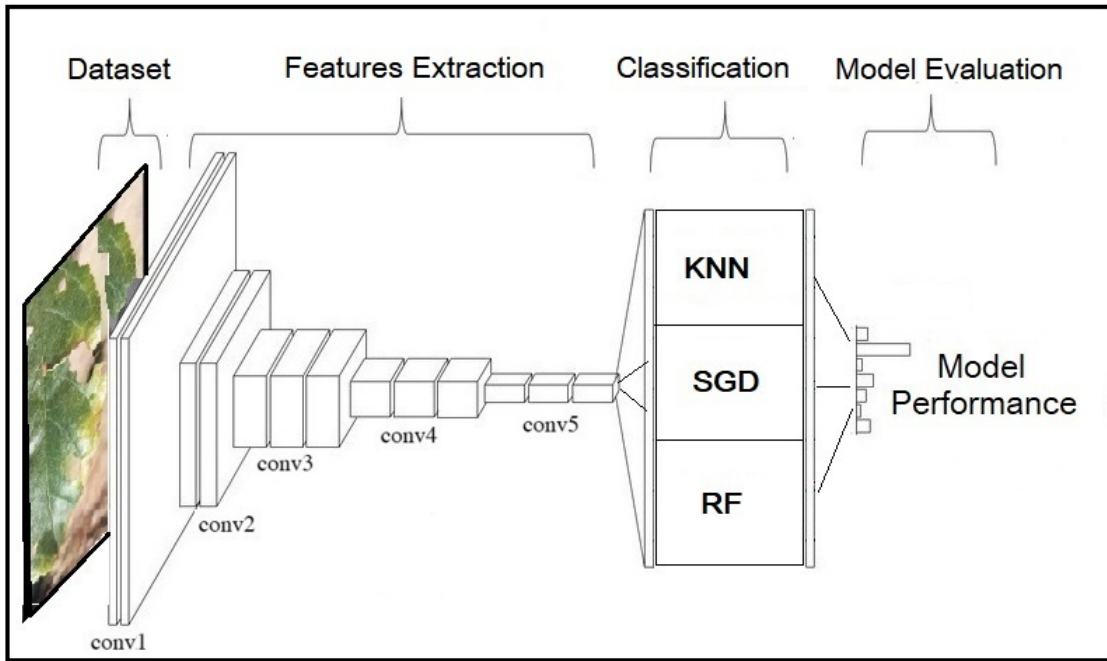


Fig. 2. An illustration of the applied research methodology, where CNN is used for extracting images' features and that are then fed to KNN, SGD, and RF classification algorithms.

$$\omega_{t+1} = \omega_t - \gamma \frac{1}{n} \sum_{i=1}^n \nabla_{\omega} Q(z_i, \omega_t) \quad (2)$$

Where  $\gamma$  denotes the model gain and  $\omega$  denotes the model weights which have to be close to optimal.

The estimated value  $z$  is randomly selected given the weight  $w_t$  for each iteration  $t$ . This process helps to optimise the resulting value in each iteration provided that the selected values are based on the ground truth distribution. The convergence of the SGD model assumes a gradual decline in the value of model gain  $\gamma$  which must be neither too fast nor too slow. The optimal value for the convergence is achieved when  $\gamma_t$  is close to the value  $t^{-1}$  which causes a decline in the error rate at the same speed [51].

SGD has been successfully used in several image classification applications. These applications involved introducing some minor modifications to improve their performance by optimising their parameters and also by combining it with other machine learning techniques such as Artificial Neural Networks (ANN) [52], [53], [54].

**Random Forests (RF):** A non-parametric, powerful machine learning technique that is used for both regression and classification [55]. The random Forests model consists of a number of regression trees that are built concurrently and then voted on selecting the best-performing model [56].

Random Forests was reported successful for image classification in several applications, which covers a wide spectrum of image processing domains [57], [58], [55]. This makes Random forests a good candidate for classifying stone fruit diseases in this study.

**3) Model Evaluation:** Cross-validation is the most widely accepted technique for evaluating the performance of machine learning classifiers [59], [60]. In this method, samples are split into several equal-size folds, and then the validation process is carried out considering one fold for testing the classification model each time, while the rest are used for training the classification model. The model validation is then repeated several times, which is equal to the number of folds. At the end of the validation process, the performance of the model is calculated by averaging the model performance for each fold.

The confusion matrix is a common evaluation method that is used to measure the performance of a machine learning model across classes. The confusion matrix is drawn in a tabular form that shows the numbers or percentage of the classified samples as illustrated in Fig. 3. It shows four types of information: (A) True positive values, which refer are the number of samples that were predicted positive and are indeed positive; (B) False-positive, which refers to the number of samples that are predicted positively by the model but they are in fact negative; (C) False-negative which refers to the number of samples that are predicted negative, but they are in fact positive, and (D) True negative which refers to the number of samples that are predicted negative and they are indeed negative.

In addition to the Classification Accuracy (CA), four other performance measures were used for evaluating the classification model: (1) Area Under the Curve (AUC); (2) F1; and (4) Precision. The confusion matrix, Receiver Operating Characteristic (ROC), Lift, and Calibration curves were also used to confirm the validity and robustness of the models performance. The ROC curve maps the false positive rate to the true positive rate of the class prediction. This curve is widely accepted as an excellent and accurate metric of the machine learning model's performance [61], [62]. In the ROC curve, the

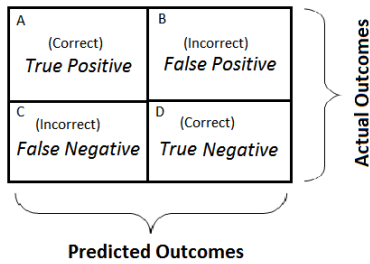


Fig. 3. An illustration of the confusion matrix. Rows show the predicted classes, while columns show the actual classes

further is a line drawn from the middle of the chart (diagonal), indicating a better performance achieved by the model. The ROC curve is also used as a basis for calculating the AUC performance metric, which is simply defined as the Area Under the ROC curve (AUC). The larger the area is, the better performance is achieved by the model [63]. The calibration curve measure the model’s performance in predicting true positive values versus false positive values, while the Lift chart measure the model’s performance in predicting positive values [64].

Classification Accuracy (CA) is one of the most commonly used performance metrics in machine learning [65]. It is used to evaluate the model’s performance in predicting the learned classes by calculating the ratio between the number of the correctly classified samples compared to the total samples, which are expressed equation 3.

$$CA = \frac{(TP + TN)}{Total} \quad (3)$$

The precision metric is also another common measure that is used for evaluating the classification model’s performance by calculating the ratio between the number of true positive values and the total number of both true positive and false positive values [65], which is expressed by equation 4.

$$Precision = \frac{TP}{(TP + FP)} \quad (4)$$

### III. RESULTS

The results obtained in this empirical study involved creating and evaluating three machine learning models using KNN, SGD, and RF algorithms. Each of these models has been created based on the feature vectors that have been extracted using the pre-trained deep learning convolutional neural networks VGG16 algorithm and using images that have been captured for stone fruit leaves that are infected with bacterial and fungal diseases.

In the first phase, the features in each image were by submitting the captured images using Keras API to a remote VGG16 algorithm implementation that is pre-trained using millions of images from the ImageNet database. Fig. 4 shows a set of images that have been generated by the VGG16 algorithm, for example, a leaf using the algorithm convolutions, pooling, and normalisation procedures. The image is a sample

of the images that are generated by each convolution in the model layers. The image was first converted into a grey-scale and was then passed to the VGG16 layers. The extracted features were stored in vectors that contain the most important descriptors for each image and saved on the local machine. The image features in the vectors were then loaded into the classifier models.

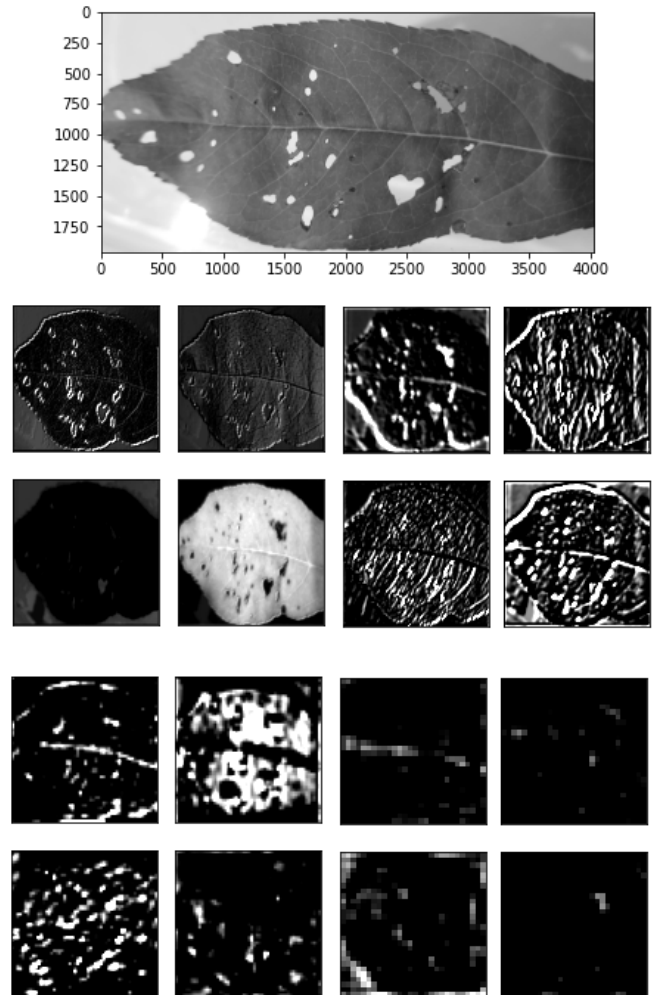


Fig. 4. Filtered image features of a leaf that suffers from fungal infection.

In the second phase, three classification models were constructed using the features extracted in the first phase. The first model was constructed using the KNN algorithm and then trained using the training dataset to predict the image classification into the two pre-labelled classes. The second model was constructed using the SGD algorithm, while the third was built using the RF algorithm.

In the third phase, the models constructed in the second phase were trained using the image dataset that was described earlier in Section II-A, which consisted of tens of images of leaves that suffer bacterial and fungal disease infections. In the fourth stage, the models were then evaluated using 10-fold cross-validation. The evaluation of each model involved measuring the performance of the classifier using: (1) Classification Accuracy; (2) Precision test; and (3) Area Under the Curve (AUC). In addition, a confusion matrix was constructed

to validate the performance of each model that was visualised using the ROC, Lift and Calibration curves [66].

The confusion matrix for the three constructed models is shown in Table IV, II, and III. The values on the diagonal of the table –highlighted with dark shadow– represent the image samples that have been correctly classified in each class, while the values to the left and the right of the table diagonal highlighted with bright shadow–, represent those images that have been wrongly classified. The values in the rows represent the actual values that, in fact, belong to the corresponding class, while the values in the columns show the values that are predicted to belong to the corresponding class.

The evaluation of the classifier models has shown good performance based on the results of the confusion matrix, ROC, Lift and calibration curves, in addition to classification accuracy and sensitivity metrics. All the applied techniques have scored more than 87.4% in classification accuracy and precision performance metrics and more than 93.9% in the AUC metric, which is an even more efficient metric for evaluating the performance of the classifiers.

The CNN-KNN outperformed the other two models in the AUC metric. It scored 98.5% in AUC, while the CNN-SGD and CNN-RF models scored 95.4% and 93.9%, respectively. On the other hand, the CNN-SGD achieved the highest score in the classification accuracy metric, which was 93.6%. In contrast, the CNN-KNN and CNN-RF models scored classification accuracy of 92.4% and 87.4%, respectively. The results of the precision and F1 support the model's scores in classification accuracy. The CNN-SGD model scored 93.7% in F1 and 94.% in Precision metrics. The CNN-KNN model scored an F1 of 92.3% and a Precision of 92.3%, while the CNN-RF scored 85.9% and 88% in F1 and Precision metrics. In the AUC metric, the CNN-KNN model scored an even higher result, just over 95.6%, while the CNN-SGD scored 89.4%. The CNN-RF model scored 88.3%. Table I compares the performance of the three models in the applied metrics.

TABLE I. A COMPARISON OF THE PERFORMANCE OF THE THREE CREATED CLASSIFICATION MODELS: CNN-KNN, CNN-SGD AND CNN-RF BASED ON AUC, ACCURACY AND PRECISION METRICS

	AUC	CA	F1	Precision
<b>CNN-KNN</b>	98.5%	92.4%	92.3%	92.3%
<b>CNN-SGD</b>	95.4%	93.6%	93.7%	94.1%
<b>CNN-RF</b>	93.9%	87.4%	85.9%	88.0%

TABLE II. THE CONFUSION MATRIX FOR THE CNN-KNN MODEL

		Predicted			Total
		Bacterial	Fungal	Healthy	
Actual	Bacterial	<b>84</b>	22	0	106
	Fungal	15	<b>279</b>	1	295
	Healthy	0	0	<b>99</b>	99
Total		99	301	100	500

The performance curves of the ROC charts also confirmed the accuracy and precision of the metric result. It showed that the CNN-KNN model by far outperformed both the SGD and RF models. These results were also apparent in predicting the fungal class and, to a lesser extent, in predicting the bacterial class. However, while the CNN-SGD model performed relatively well according to the ROC chart, the performance of

TABLE III. THE CONFUSION MATRIX FOR THE CNN-SGM MODEL

		Predicted			Total
		Bacterial	Fungal	Healthy	
Actual	Bacterial	<b>97</b>	8	1	106
	Fungal	23	<b>272</b>	0	295
	Healthy	0	0	<b>99</b>	99
Total		120	280	100	500

TABLE IV. THE CONFUSION MATRIX FOR THE CNN-RF MODEL

		Predicted			Total
		Bacterial	Fungal	Healthy	
Actual	Bacterial	<b>49</b>	57	0	106
	Fungal	6	<b>289</b>	0	295
	Healthy	0	0	<b>99</b>	99
Total		55	346	99	500

the RF model lagged. Fig. 5 illustrate the performance of the three constructed models using the ROC chart. Furthermore, a calibration curve was created for all the constructed models, which compares them in classifying true and false positive classes. The threshold of the P value was set to 0.50, which is shown in Fig. 6. The SGD classifier scored a rate of 91.5% and 10% in predicting true positive and 8.8% in predicting false positive samples. The KNN classifier scored 79.2% and 3.8% in predicting the true positive and false positive samples, respectively, while the Random Forests model scored 39.6% and 10% in predicting the true positive and false positive samples, respectively. These results, yet again, show that both the SGD and KNN significantly outperformed the Random Forest model. However, the SGD model achieved the best performance, particularly in predicting true positive values. The Lift curve in Fig. 7 shows the performance of the three classifiers in predicting the positive values (Positive rate). Based on the Lift curve, the SGD model achieved the best performance, followed by the KNN and SGD models.

The champion two models were constructed by combining the pre-trained convolutional network algorithm VGG16 with the KNN and SGD algorithms. The CNN-KNN model scored 98.5% in the AUC metric, while the CNN-SGD scored a classification accuracy of 93.6%. However, the AUC metric is considered by many data mining practitioners as the best performance metric for evaluating machine learning models [63].

#### IV. DISCUSSION

This research involved collecting and photographing hundreds of leaves from stone fruit trees that are grown in the north of Jordan that suffer from bacterial and fungal diseases. The leaves were photographed and then classified based on their identified infections. The images were then fed into a deep learning convolutional network neural network algorithm that was used to identify the significant features involved in each of them. The resulting features were then used as input for three machine learning algorithms that have been used for classifying the leaves' images according to their infected diseases. The constructed models were then trained, and their performances were measured and evaluated based on a 10-fold cross validation.

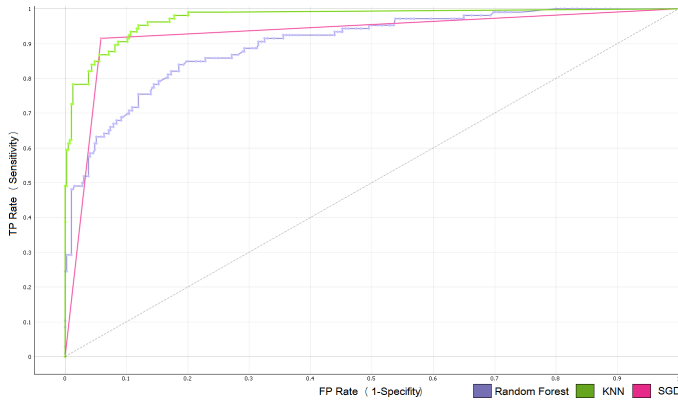


Fig. 5. An ROC chart that compares the performance of the three created classification models.

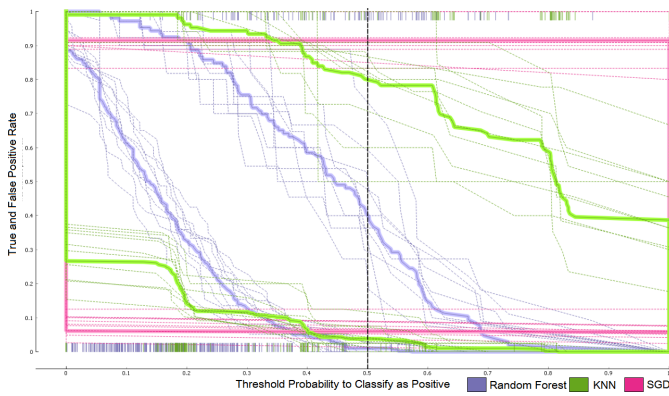


Fig. 6. A Calibration curve showing the performance of the three created classification models in predicting true positive and false positive classes

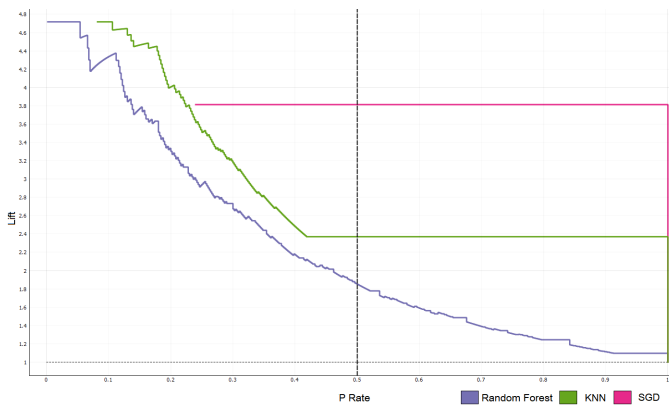


Fig. 7. A Lift curve showing the performance of the three created classification models in terms of the relationship between lift and the classification positive rate

The results confirmed the research hypothesis, which suggested that pre-trained deep learning algorithms can provide an efficient method for diagnosing stone fruit fungal and bacterial diseases by identifying image features and then using them in classifying the diseases' images using machine learning classification algorithms.

The results also provides a positive answer to the research question by creating classification models that were able to

diagnose the stone fruit fungal and bacterial infections with an accuracy performance of 93.6% using the CNN-SGD model and an AUC performance of 98.5% using the CNN-KNN model.

The study was successful in achieving the preset research objectives, which aimed at creating a computational model that would help create an automated prediction model that can be embedded in a mobile phone application or a web service to provide a fast, cheap, and reliable diagnosis of fungal and bacterial infections through photographing the visual symptoms that on plant leaves using standard mobile cameras and under the field's normal conditions. The findings of the research confirm the validity of our proposed approach to achieve the research objectives through combining CNN algorithm for feature extraction with three classical machine learning classifiers: KNN, SGD and Random Forest.

The performance of the classification model created in this study was above the average performance of other models that have been reported in similar and previous studies, particularly when compared to other studies that involved collecting, photographing, and analysing hundreds of plant leaf images under normal field conditions [22], [23], [25], [26], [27]. However, other studies that used public datasets with controlled conditions and larger dataset achieved better results as reported in [29], [30], [31], [32], [35].

The CNN-KNN model outperformed both CNN-SGD and CNN-RF in the AUC metric, as it scored a performance of 98.5%, while they scored an AUC performance of 95.4% and 93.9%, respectively. On the other hand, the CNN-SGD model outperformed the other two models based on the classification accuracy metric as its scored performance of 93.6%. The CNN-KNN and CNN-RF models scored a classification accuracy of 92.8% and 87.4%, respectively. However, most of the other studies that reported better results used thousands of images from public benchmark datasets that were acquired under controlled environments and conditions. In contrast, our study did not suffer some of the problems that are related to the degradation of model performance when leaves were taken on a different background, as reported in [30].

The models created in the present study is quite rigorous, even when images were taken in different aspect ratios and at different orientations. In addition, most of the studies that we have referred to in the related work section lack consideration for the prospect of the acquired knowledge and for the practical deployment and realisation of the generated model, which was considered from the beginning and to the end of this study.

## V. CONCLUSION

In this research, the power of deep learning in feature extraction and detection was combined with the classification power of three machine learning algorithms: KNN, SGD, and RF, for analysing leaf infections visually using image recognition and classification technology.

The study concerned conducting end-to-end empirical research, which involved collecting hundreds of stone fruit leaf samples from fields that suffer from bacterial and fungal infections. The samples were collected and then photographed from farms in various locations in the north of Jordan. The

bacterial and fungal infections were diagnosed and identified by two experts in plant diseases and confirmed using literature and using the European and Mediterranean Plant Protection Organisation (EPPO) database repository, which covers 1,700 pests and 88,000 plant species.

The results confirm the applicability and validity of combining the deep learning CNN algorithm with other machine learning classification algorithms for identifying and diagnosing plant bacterial and fungal diseases based on their visual signs and symptoms. The CNN-KNN model outperformed both the CNN-SGD and CNN-RF models based on the AUC metric and ROC curve, as it scored a performance of 98.5% in AUC. On the other hand, the CNN-SGD model's performance was the best in the classification accuracy metric, scoring 93.6%. The CNN-KNN model scored 92.4%, while the CNN-RF model scored 87.4%. Nevertheless, the limitations of this study come from its specialisation in one crop rather than others, as well as in its limited coverage of bacterial and fungal infections rather than other diseases and deficiencies. Nevertheless, this limitation was noticed in most of the previous works, including those that used public and field datasets. Expanding the research to cover more plants and more diseases would provide a more comprehensive image of other crops and other diseases.

The future work of this research may involve conducting further research to cover a wider range of crops and diseases under the field's normal conditions and using thousands of images. It may also involve using more deep learning methods and machine learning algorithms to achieve better results. In addition, constructing a data repository for all the diagnosed diseases would also help in enhancing the accuracy of future models by providing more images for model training and testing.

Furthermore, the success of these applications can be utilised in deploying machine learning classification models that can be embedded in a mobile application that can provide fast, cheap, and reliable identification of plant diseases or can be integrated into a web-based service that can help farmers across the globe to diagnose their plant's diseases by photographing the leaf infections in their crops using their mobile phones and upload the images to the web service.

#### ACKNOWLEDGEMENT

We would like to thank Mr. Abu Hmiadan and other anonymous farmers who facilitated the collection of samples.

#### REFERENCES

- [1] V. A. Gulhane and A. A. Gurjar, "Detection of diseases on cotton leaves and its possible diagnosis," *International Journal of Image Processing (IJIP)*, vol. 5, no. 5, pp. 590–598, 2011.
- [2] J. Adaskaveg, J. Ogawa, and E. Butler, "Morphology and ontogeny of conidia in wilsonomyces carpophilus, gen. nov. and comb. nov., causal pathogen of shot hole disease of prunus species," *Mycotax*, vol. 37, pp. 275–290, 1990.
- [3] C. Văcăroju, C. Zală, and S. Cristea, "Research on the morphology and biology of the stigmata carpophila fungus," in *scientific conferences with international participation durable agriculture-Agriculture of future, 4th edition. Craiova (Romania): Faculty of Agriculture, University of Craiova*, vol. 110, 2008.
- [4] J. M. Ogawa and H. English, *Diseases of temperate zone tree fruit and nut crops*. UCANR Publications, 1991, vol. 3345.

- [5] A. Ahmadpour, Y. Ghosta, M. Javan-Nikkhah, K. Ghazanfari, and R. Fatahi, "Study on morphology, pathogenicity and genetic diversity of wilsonomyces carpophilus isolates, the causal agent of shot hole of stone fruit trees based on rapid-pcr in iran," *Archives of Phytopathology and Protection, Plant*, vol. 45, no. 17, pp. 2076–2086, 2012.
- [6] H. Khlaf, I. Abu-Obeid, and B. Werikat, "Occurrence of plant bacterial diseases in Jordan," *African Journal of Agricultural Research*, vol. 13, no. 40, pp. 2104–2117, 2018.
- [7] L. Highberg and J. Ogawa, "Survival of shot-hole inoculum in association with dormant almond buds," *Journal of Plant Disease*, vol. 70, no. 9, pp. 828–831, 1986.
- [8] L. Gianessi and N. Reigner, "The value of fungicides in us crop production. croplife foundation," *Journal of Crop Protection Research Institute*, 2005.
- [9] D. Ritchie, "Sprays for control of bacterial spot of peach cultivars having different levels of disease susceptibility," *Fungicide Nematicide Tests*, vol. 54, p. 63, 1999.
- [10] E. Alpaydin, *Machine learning*. MIT Press, 2021.
- [11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [12] G. Hinton, "Deep learning a technology with the potential to transform health care," *Jama*, vol. 320, no. 11, pp. 1101–1102, 2018.
- [13] A. BaniMustafa, "Predicting software effort estimation using machine learning techniques," in *2018 8th International Conference on Computer Science and Information Technology (CSIT)*. IEEE, 2018, pp. 249–256.
- [14] M. Pak and S. Kim, "A review of deep learning in image recognition," in *2017 4th international conference on computer applications and information processing technology (CAIPT)*. IEEE, 2017, pp. 1–3.
- [15] R. Gajjar, N. Gajjar, V. J. Thakor, N. P. Patel, and S. Ruparelia, "Real-time detection and identification of plant leaf diseases using convolutional neural networks on an embedded platform," *The Visual Computer*, vol. 38, no. 8, pp. 2923–2938, 2022.
- [16] M. M. Kabir, A. Q. Ohi, and M. F. Mridha, "A multi-plant disease diagnosis method using convolutional neural network," in *Computer Vision and Machine Learning in Agriculture*. Springer, 2021, pp. 99–111.
- [17] A. BaniMustafa, "Enhancing learning from imbalanced classes via data preprocessing a data-driven application in metabolomics data mining," *The ISC International Journal of Information Security*, vol. 11, no. 3, pp. 79–89, 2019.
- [18] A. BaniMustafa and N. Hardy, "Applications of a novel knowledge discovery and data mining process model for metabolomics," *arXiv preprint arXiv:1907.03755*, 2019.
- [19] A. BaniMustafa, M. Baklizi, and K. Khatatneh, "Machine learning for securing traffic in computer networks," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 12, pp. 426–435, 2022.
- [20] M. H. Sheikh, T. T. Mim, M. S. Reza, A. S. A. Rabby, and S. A. Hossain, "Detection of maize and peach leaf diseases using image processing," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2019, pp. 1–7.
- [21] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and electronics in agriculture*, vol. 147, pp. 70–90, 2018.
- [22] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," *IEEE Access*, vol. 7, pp. 59 069–59 080, 2019.
- [23] X. Xie, Y. Ma, B. Liu, J. He, S. Li, and H. Wang, "A deep-learning-based real-time detector for grape leaf diseases using improved convolutional neural networks," *Frontiers in plant science*, vol. 11, 2020.
- [24] E. Fujita, Y. Kawasaki, H. Uga, S. Kagiwada, and H. Iyatomi, "Basic investigation on a robust and practical plant diagnostic system," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2016, pp. 989–992.
- [25] D. Oppenheim and G. Shani, "Potato disease classification using convolution neural networks," *Advances in Animal Biosciences*, vol. 8, no. 2, p. 244, 2017.

- [26] J. Chen, Q. Liu, and L. Gao, "Visual tea leaf disease recognition using a convolutional neural network model," *Symmetry*, vol. 11, no. 3, p. 343, 2019.
- [27] J. Amara, B. Bouaziz, and A. Algergawy, "A deep learning-based approach for banana leaf diseases classification," in *Datenbanksysteme for Business, Technologie und Web (BTW 2017) - Workshopband*, B. Mitschang, D. Nicklas, F. Leymann, H. Schning, M. Herschel, J. Teubner, T. Hörder, O. Kopp, and M. Wieland, Eds. Bonn: Gesellschaft für Informatik e.V., 2017, pp. 79–88.
- [28] K. Yamamoto, T. Togami, and N. Yamaguchi, "Super-resolution of plant disease images for the acceleration of image-based phenotyping and vigor diagnosis in agriculture," *Sensors*, vol. 17, no. 11, p. 2557, 2017.
- [29] S. Wallelign, M. Polceanu, and C. Buche, "Soybean plant disease identification using convolutional neural network," in *The thirty-first international flairs conference*, 2018.
- [30] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in plant science*, vol. 7, p. 1419, 2016.
- [31] P. V. Raja, K. Sangeetha, N. B. A., S. M., and S. S. S., "Convolutional neural networks based classification and detection of plant disease," in *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, 2022, pp. 1484–1488.
- [32] M. Pei, M. Kong, M. Fu, X. Zhou, Z. Li, and J. Xu, "Application research of plant leaf pests and diseases base on unsupervised learning," in *2022 3rd International Conference on Computer Vision, Image and Deep Learning and International Conference on Computer Engineering and Applications (CVIDL and ICCEA)*, 2022, pp. 1–4.
- [33] H. Wang, S. Shang, D. Wang, X. He, K. Feng, and H. Zhu, "Plant disease detection and classification method based on the optimized lightweight yolov5 model," *Agriculture*, vol. 12, no. 7, p. 931, 2022.
- [34] Z. Liu, R. N. Bashir, S. Iqbal, M. M. A. Shahid, M. Tausif, and Q. Umer, "Internet of things (iot) and machine learning model of plant disease prediction–blister blight for tea plant," *IEEE Access*, vol. 10, pp. 44 934–44 944, 2022.
- [35] P. Praveen, M. Nischitha, C. Supriya, M. Yogitha, and A. Suryanandh, "To detect plant disease identification on leaf using machine learning algorithms," in *Intelligent System Design*. Springer, 2023, pp. 239–249.
- [36] 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*. IEEE, 2009, pp. 248–255.
- [37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [38] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge, 2014," *arXiv preprint arXiv:1409.0575*, 2014.
- [39] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.
- [40] A. Gulli and S. Pal, *Deep learning with Keras*. Packt Publishing, 2017.
- [41] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019.
- [42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
- [43] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.
- [44] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
- [45] A. BaniMustafa and N. Hardy, "A strategy for selecting data mining techniques in metabolomics," in *Plant Metabolomics: Methods and Protocols*, ser. Methods in Molecular Biology, N. Hardy and R. Hall, Eds. Springer Science, 2012, vol. 860, ch. 18, pp. 317–333.
- [46] A. Banimustafa and N. Hardy, "A scientific knowledge discovery and data mining process model for metabolomics," *IEEE Access*, vol. 8, pp. 209 964–210 005, 2020.
- [47] A. M. Bani Mustafa, "A knowledge discovery and data mining process model for metabolomics," Ph.D. dissertation, Aberystwyth University, 2012.
- [48] E. Fix, *Discriminatory analysis: nonparametric discrimination, consistency properties*. USAF School of Aviation Medicine, 1951.
- [49] R. Ramteke and K. Y. Monali, "Automatic medical image classification and abnormality detection using k-nearest neighbour," *International Journal of Advanced Computer Research*, vol. 2, no. 4, p. 190, 2012.
- [50] J. Kim, B.-S. Kim, and S. Savarese, "Comparing image classification methods: K-nearest-neighbor and support-vector-machines," in *Proceedings of the 6th WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics*, 2012, pp. 133–138.
- [51] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.
- [52] A. El Mouatasim, "Fast gradient descent algorithm for image classification with neural networks," *SIGNAL IMAGE AND VIDEO PROCESSING*, 2020.
- [53] B. Zhang, C. Chen, Q. Ye, J. Liu, D. Doermann *et al.*, "Calibrated stochastic gradient descent for convolutional neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 9348–9355.
- [54] Z. Yang, C. Wang, Z. Zhang, and J. Li, "Accelerated stochastic gradient descent with step size selection rules," *Signal Processing*, vol. 159, pp. 171–186, 2019.
- [55] N. Horning *et al.*, "Random forests: An algorithm for image classification and generation of continuous fields data sets," in *Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Osaka, Japan*, vol. 911, 2010.
- [56] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [57] B. Xu, Y. Ye, and L. Nie, "An improved random forest classifier for image classification," in *2012 IEEE International Conference on Information and Automation*. IEEE, 2012, pp. 795–800.
- [58] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *2007 IEEE 11th international conference on computer vision*. Ieee, 2007, pp. 1–8.
- [59] M. W. Browne, "Cross-validation methods," *Journal of mathematical psychology*, vol. 44, no. 1, pp. 108–132, 2000.
- [60] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14. Montreal, Canada, 1995, pp. 1137–1145.
- [61] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
- [62] J. Fan, S. Upadhye, and A. Worster, "Understanding receiver operating characteristic (roc) curves," *Canadian Journal of Emergency Medicine*, vol. 8, no. 1, pp. 19–20, 2006.
- [63] C. X. Ling, J. Huang, H. Zhang *et al.*, "Auc: a statistically consistent and more discriminating measure than accuracy," in *Ijcai*, vol. 3, 2003, pp. 519–524.
- [64] M. Vuk and T. Curk, "Roc curve, lift chart and calibration plot," *Advances in methodology and Statistics*, vol. 3, no. 1, pp. 89–108, 2006.
- [65] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management*, vol. 45, no. 4, pp. 427–437, 2009.
- [66] J. D. Novaković, A. Veljović, S. S. Ilić, Ž. Papić, and M. Tomović, "Evaluation of classification models in machine learning," *Theory and Applications of Mathematics & Computer Science*, vol. 7, no. 1, pp. 39–46, 2017.

# Enhancing Collaborative Interaction with the Augmentation of Sign Language for the Vocally Challenged

Sukruth G L, Dr. Vijaya Kumar B P, Tejas M R, Rithvik K, Trisha Ann Tharakan  
Department of Information Science and Engineering, M S Ramaiah Institute of Technology  
Bangalore, Karnataka, India

**Abstract**—As per Census 2011, in India, there were 26.8 million differently abled people, out of which more than 25% of the people faced difficulty in vocal communication. They use Indian Sign Language (ISL) to communicate with others. The proposed solution is developing a sensor-based Hand Gesture Recognition (HGR) wearable device capable of translating and conveying messages from the vocally challenged community. The proposed method involves designing the hand glove by integrating flex and Inertial Measurement Unit (IMU) sensors within the HGR wearable device, wherein the hand and finger movements are captured as gestures. They are mapped to the ISL dictionary using machine learning techniques that learn the spatio-temporal variations in the gestures for classification. The novelty of the work is to enhance the capacity of HGR by extracting the spatio-temporal variations of the individual's gestures and adapt it to their dynamics with aging and context factors by proposing Dynamic Spatio-temporal Warping (DSTW) technique along with long short term memory based learning model. Using the sequence of identified gestures along with their ISL mapping, grammatically correct sentences are constructed using transformer-based Natural Language Processing (NLP) models. Later, the sentences are conveyed to the user through a suitable communicable media, such as, text-to-voice, text-image, etc. Implementation of the proposed HGR device along with the Bidirectional Long-Short Memory (BiLSTM) and DSTW techniques is carried out to evaluate the performance with respect to accuracy, precision and reliability for gesture recognition. Experiments were carried out to capture the varied gestures and their recognition, and an accuracy of 98.91% was observed.

**Keywords**—*Hand Gesture Recognition (HGR); wearable sensors; Long-Short Term Memory (LSTM); Natural Language Processing (NLP); Dynamic Spatio-Temporal Warping (DSTW); Indian Sign Language (ISL)*

## I. INTRODUCTION

The need for communication stems from our natural desire to express ourselves. In a world that is becoming more inclusive, no one should feel left behind. The differently-abled community has been struggling historically to assimilate with the society. There exist natural communication barriers between the differently abled and abled people. This is because, the differently abled community, generally the vocally challenged community, uses sign language to communicate with others. Around 99% of the world do not understand sign language [1]. This creates a massive communication gap between the people in such communities who only use sign language for communication and those who do not understand sign language. The traditional solution for this has been to use an interpreter, who specializes in both sign language and

spoken language. With the advent of technology, especially in the fields of artificial intelligence and machine learning, a hand gesture recognition system for sign language can be developed which can replace the interpreter, making the people of the differently abled people independent.

Sign languages (also known as signed languages) are languages that use the visual-manual modality to convey meaning. In India, the primary sign language used is the Indian Sign Language (ISL). ISL is used in the deaf and/or dumb community all over India.

ISL interpreters are an urgent requirement at institutes and places where communication between deaf and hearing people takes place, but India has around 300 certified interpreters, which is minuscule to the number of interpreters that are required [2]. The needs of the deaf and/or dumb community have long been ignored, and the problems have been documented by various organizations working for them. The aim is to bridge this gap with the help of a sign language interpreter by using HGR.

This paper proposes the use of sensor-based gesture recognition where sensors are placed on each finger and on the hand, through which the sensor data is collected based on the hand and finger movements. Such raw data is transformed to gesture data and classified using machine learning algorithms. The obtained results on classification of sequence of gestures is fed to the NLP models involving segmentation and grammar correction. Finally, the sentences are conveyed suitably through communicable media devices like speakers, displays, braille tablets, etc.

The remaining paper has been structured as follows. Section II discusses some related works done and the key points from different research papers. Section III elaborates on the proposed model for sign language gesture recognition. Section IV explains the implementation methodologies. Section V describes along with the inferences for the proposed model and results with respect to performance parameters. Section VI provides the concluding remarks and future scope.

## II. RELATED WORK

Some of the related works in the areas of hand gesture recognition and NLP related to differently abled communication are abstracted and discussed in the section.

Flex sensors are used to develop a HGR based system by capturing the finger movements by making use of the

Flex ADC values, voltage, resistance, and the ratio between the flex voltage and source voltage values. These inputs are then processed through a GRU and classification is done by implementing maximum a posteriori [3].

Another methodology wherein a yarn based stretchable sensor array (YSSAs) was made use of, which is based on the change in voltage as and when a gesture is being performed [4]. It is noticed here that the YSSAs, is not a commercially available product. The Flex or the IMU sensors are available commercially and at a low cost, which would provide ease of access to the materials of the data glove.

In a system of sensor-based HGR by making use of accelerators and gyroscopes, the most important part is the determination of the starting and stopping points whilst performing the continuous hand gesture and it was solved by implementing an LSTM model in [5]. Based on a many-to-many interface scheme, where output is produced by the LSTM at each time step, a collection of the output sequence is viewed as an output path.

To tackle long-distance dependencies, a frame stream density compression (FSDC) algorithm is introduced in [6] for detecting and reducing redundant similar frames. The traditional encoder is replaced in a neural machine translation (NMT) module with an improved architecture, which incorporates a temporal convolution (T-Conv) unit and a dynamic hierarchical bidirectional GRU (DH-BiGRU) unit sequentially.

It is often necessary in practical situations to attempt parsing an incorrect or incomplete input. Since there are no signs for articles (i.e., the, is, an, etc.) most sentences performed using ISL will be structurally and grammatically incorrect. To tackle this issue certain NLP solutions were explored, one of them was parsing incomplete sentences. A chart parser was used in [7] to try all possible parses on all possible inputs. The result is a parse forest for all the grammatically acceptable sentences that can be generated by the (non-necessarily deterministic) finite state automaton.

Recent work on Grammar Error Correction (GEC) has highlighted the importance of language modeling to improve the performance and also compared the probabilities of the proposed edits. The approach in [8] involves using Google's BERT and OpenAI's Generative Pre-Trained Transformer (GPT) and GPT-2. Using simple heuristics, the language model generates a score such that it reduces the error and confusion set of sentences.

A more efficient algorithm using the distant measurement scheme of Dynamic Time Warping (DTW) and learning method of Restricted Coulomb Energy (RCE) neural network is proposed in [9]. A test platform is constructed using an IMU sensor to verify that real-time learning and recognition was possible using the proposed algorithm. DTW-based acknowledgment calculations have shown great execution for HGR.

Inferencing from the above literature review, using sensors and a hand glove design is provided with a sensor-based data collection system to capture finger and hand movements, a gesture classification unit along with dynamic spatio-temporal warping (DSTW) algorithm, grammar correction unit and a text to communicable media unit.

The data collected from the glove is in the form of time-series data, which is a type of Sequential Data. Sequential neural networks are designed to classify sequential data. Recurrent neural networks have the vanishing and exploding gradient problems, which is handled by LSTM by introducing cell state to decide what information to retain and forget long term. BiLSTM is used because it has shown better accuracy than LSTM models.

### III. PROPOSED MODEL

In this work, the novelty involves in designing a suitable sensor-based glove that consists of flex and IMU sensors to collect the hand and finger movements as signs and gestures. Here, the glove is calibrated for different dynamics of hand and finger movements collected as time-series data. Such time-series data is analyzed for spatio-temporal variations using machine learning algorithms that use neuro-computing models. The different spatio-temporal variations are mapped to the respective gestures taken from ISL dictionary [10]. However, to capture similar gestures performed by individuals that vary w.r.t time and space, the dynamic spatio-temporal warping technique is adopted. To improve the accuracy and generalization, different gestures with time and space variations with continuous learning for classification and repeated mapping to the ISL dictionary can be performed to avoid over-fitting and under-fitting. From the trained model, a given series of gestures is converted to a meaningful sentence using a transformer-based Natural Language Processing (NLP) model. Such sentences can be converted into the text/speech or any other media that an abled human can understand.

Here, the learning methods include feed forward neural network model with memory-based learning integrated with DSTW as similarity metric along with a BiLSTM neural network model with Online Learning. Gestures belonging to the same category which are partially and semantically similar can be classified accurately even if there are time and space variations in any given gesture, as the experience increases with learning. The basic idea of training using memory-based learning is that it classifies the different gestures and also their similarities with the previously seen gesture data.

The ability of the proposed system is to compute and abstract the spatio-temporal variations in the data items derived from hand and finger movements, as time-series data items and their correlations using feed-forward neuro-computing models.

In time-series data analysis, dynamic spatio-temporal warping is an algorithm for measuring similarity between two temporal sequences, which may vary in time and space. For instance, similarities in walking could be detected using DSTW, even if one person is walking faster than the other, or if there is acceleration and/or deceleration during the course of an action. DSTW is illustrated in Section IVB with mathematical formulation.

The complete proposed model and its functional modules is shown in Fig. 1. The sensors data corresponding to a gesture taken from ISL video dictionary [10], is collected from the hand glove and it is pre-processed to extract time and spatial features for its respective meaning. Such mapping is continuously fed into neuro-computing learning modules until the gesture meanings are mapped correspondingly. However,



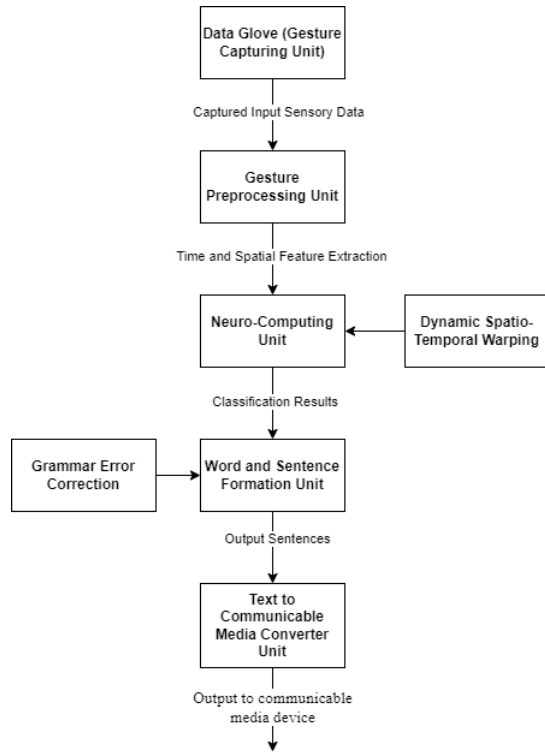


Fig. 1. Functional modules of the proposed hand gesture recognition (HGR) system.

the speed and area of occupancy in performing the same gestures by different individuals will vary, hence, DSTW algorithm is integrated to classify and map the meaning to the ISL dictionary. The sequence of gestures performed are integrated to generate meanings by using transformer-based NLP model with grammar correction to obtain a grammatically correct sentence. Such generated sentences are converted into communicable media by using text-to-speech, text-to-images, etc. as per the user's requirement.

#### A. Moving Average Smoothing

The captured data for the performed gestures will be in the form of time-series data and is bound to be noisy. To eliminate the fine-grained fluctuation between time steps, smoothing is used over the time-series data. Here, the smoothing is intended to reduce noise and thus clearly reveal the signal of underlying causal processes. For time-series data, moving averages are a straightforward and widely used method of smoothing. Moving Average Smoothing, represented in Fig. 2, involves creating a new time-series where the values are the average of raw observations in the original time-series.

A window size, referred to as the window width, is necessary when using a moving average. This specifies how many unprocessed observations were utilized to determine the moving average value. In order to calculate the average values in the new series, the window defined by the window width is moved along the time-series, hence giving the name "moving average".

After the noise has been removed through the moving

average smoothing process, the time and spatial features (i.e. the bend angle of flex sensors and the Roll, Pitch and Yaw of the IMU Sensor) is fed to the neuro-computing unit, wherein the preprocessed data is fed to various machine learning and neural network models for the classification of the gestures performed.

#### B. Dynamic Spatio-Temporal Warping (DSTW)

This subsection discusses the dynamic spatio-temporal warping algorithm used for comparing the ISL gesture data collected through the glove, consisting of flex and IMU sensors placed suitably to abstract the clear movement of fingers and hand.

DSTW uses time and space variations as that of time in DTW [11], [12]. The DSTW algorithm and how it can be applied to handle time and space variations of two similar gestures and their analysis is explained below.

Considering two data sequences belonging to gestures  $S$  and  $T$  of varying time lengths  $N$  and  $M$  samples, which are sampled at the same rate (i.e. the time taken to perform the two gestures  $S$  and  $T$  may not be the same). Let:

$$S = (S_1, S_2, \dots, S_N) \text{ and } T = (T_1, T_2, \dots, T_M) \quad (1)$$

be the measured values at the sampling times  $t_1, t_2, \dots, t_N$  &  $t_1, t_2, \dots, t_M$  for  $S$  and  $T$ , respectively.

In order to find the similarities between the two sequences, a cost matrix  $C$  of  $N \times M$  dimension is defined and is formulated using the following equations:

$$C(0, 0) = 0 \quad (2)$$

$$C(i, 0) = \infty \quad \forall i \in [1, N] \quad (3)$$

$$C(0, j) = \infty \quad \forall j \in [1, M] \quad (4)$$

$$C(i, j) = |S_i - T_j| + \min(C(i-1, j), C(i, j-1), C(i-1, j-1)) \quad \forall i \in [1, N] \text{ and } j \in [1, M] \quad (5)$$

With this cost matrix, the goal is to find the optimal path, which has the minimal overall cost that leads to gestures  $S$  and  $T$  having similarities to the maximum extent as compared to dissimilar gestures, as per the gestures given in the ISL video dictionary.

In order to have accuracy in the similarity measure is to follow certain necessary conditions i.e. to find the optimal path by traversing the minimum value in the cost matrix involving the time series gesture data formulated in eq. (5) are: (i) To travel the cost matrix from top left corner to bottom right corner, (ii) the path should be incremental in steps, (iii) the path should move from one cell  $(i, j)$  at a time, either to the right  $(i+1, j)$  or bottom  $(i, j+1)$  or bottom-right  $(i+1, j+1)$  cell by choosing the cell with minimum of the costs  $C(i+1, j)$ ,  $C(i, j+1)$ ,  $C(i+1, j+1)$  respectively as represented by equations 8, 9, 10, respectively.

Here, the optimal path  $O$  with length  $L$  can be obtained by traversing the minimum value of the cost matrix, is as follows:

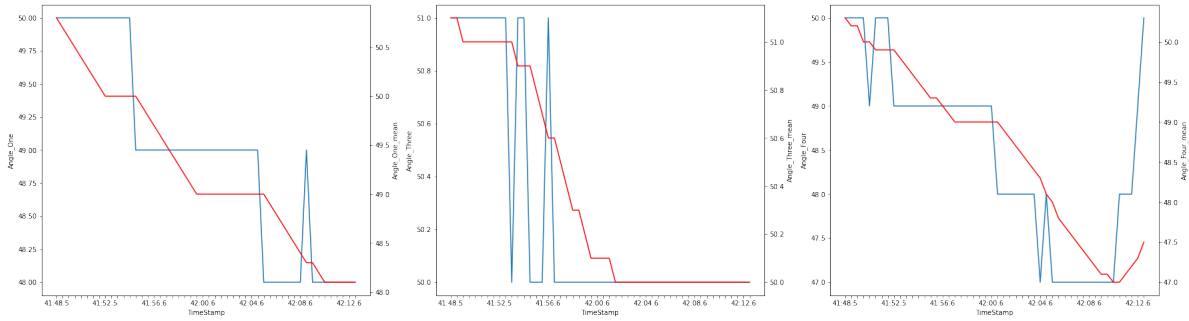


Fig. 2. Moving average implementation on gestures captured by data glove.

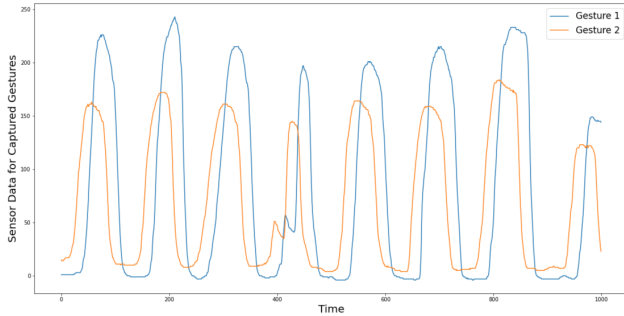


Fig. 3. Comparing two similar gestures of varying lengths with DSTW.

$$O = (O_1, \dots, O_l, \dots, O_L) \quad (6)$$

$$O_l = (i, j) \text{ for some } i \in [1, N] \text{ and } j \in [1, M] \quad (7)$$

$$O_1 = (1, 1) \text{ and } O_L = (N, M) \quad (8)$$

$$O_1 \leq O_2 \leq \dots \leq O_{L-1} \leq O_L \quad (9)$$

$$(O_{l+1} - O_l) \in (1, 0), (0, 1), (1, 1) \text{ for } l \in [1, L - 1] \quad (10)$$

The time series data discussed above and their respective mapping in space and time domain along with their sampled values is illustrated with examples below.

Two gestures of the same kind having different time and space variations can be compared using DSTW. Consider two gestures of the same kind,  $G = [1, 2, 3, 4, 5]$  and  $G' = [1, 1, 2, 2, 3, 3, 4, 4, 5, 5]$ . Both the gestures are the same but the first gesture is being performed twice as fast as the second one. The same is the case with space variations. Consider two gestures of the same kind  $G = [2, 2, 3, 3, 3]$  and  $G' = [3, 3, 4, 4, 4]$ . These gestures are varying with respect to space, which means the magnitude of the gestures being performed is different, but the gesture remains the same.

In order to handle the above situation, DSTW Algorithm 1 is proposed to integrate time and space for comparing the gestures carried out with different time durations to handle both the intra-sampling and inter-sampling variations. Comparison of two gestures using DSTW algorithm is shown in Fig. 3. The output of Algorithm 1 is the distance between the two gestures  $G$  and  $G'$ .

### C. BiLSTM Architecture

Similarly, the spatio-temporal variations with similarity principles are considered intrinsically in BiLSTM architecture and is explained below. BiLSTM architecture for classifying the gestures, as shown in Fig. 4, consists of two LSTM cells; forward LSTM and backward LSTM [13]. The forward LSTM processes the time-series gesture information from left to right and its hidden state  $\vec{h}$  can be shown as

$$\vec{h} = LSTM(x_t, \vec{h}_{t-1}) \quad (11)$$

where  $x_t$  is the time series gesture data point at time  $t$  and  $\vec{h}_i$  is the hidden state of the forward LSTM at time  $t = i$ .

The backward LSTM processes the time-series gesture information from right to left and its hidden state  $\overleftarrow{h}$  can be expressed as

$$\overleftarrow{h} = LSTM(x_t, \overleftarrow{h}_{t+1}) \quad (12)$$

where  $\overleftarrow{h}_i$  is the hidden state of the backward LSTM at time  $t = i$ .

Finally, the output of BiLSTM can be summarized by concatenating the forward and backward states as

$$h_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (13)$$

Correlation of  $\vec{h}_t$  and  $\overleftarrow{h}_t$  at time  $t$  is performed to get better classification and feature extraction, since gestures will have spatio-temporal variations with redundant and varying features in different time spans.

In order to bring better accuracy of feature extraction in gestures, online learning is proposed. Online learning represented in Fig. 5 is done by performing the training process with one data point at a time. Fig. 5 shows the training process of the model where  $G_i$  is processed at time  $t = i$ . This approach is common when working with sequential data.

Once these gestures are classified and their meanings are mapped using ISL, then the obtained meanings are fed sequentially into the word and sentence formation unit with transformer-based NLP model to form grammatically correct sentences.

**Algorithm 1:** Dynamic Spatio-Temporal Warping

**Input:** Collected data for two gestures  $G = [g_1, \dots, g_N]$ , and  $G' = [g'_1, \dots, g'_M]$  as arrays of varying lengths  $N$  and  $M$   
**Output:** Similarity measure between  $G$  and  $G'$   
 $SpaceTimeGap \leftarrow matrix[0 \dots N, 0 \dots M]$   
for  $i \leftarrow 0$  to  $N$  do  
  for  $j \leftarrow 0$  to  $M$  do  
     $SpaceTimeGap[i, j] \leftarrow \infty$   
 $SpaceTimeGap[0, 0] \leftarrow 0$   
for  $i \leftarrow 1$  to  $N$  do  
  for  $j \leftarrow 1$  to  $M$  do  
     $cost \leftarrow distanceMeasure(g_i, g'_j)$   
     $SpaceTimeGap[i, j] \leftarrow cost + \min(SpaceTimeGap[i - 1, j], SpaceTimeGap[i, j - 1], SpaceTimeGap[i - 1, j - 1])$   
return  $SpaceTimeGap[N, M]$

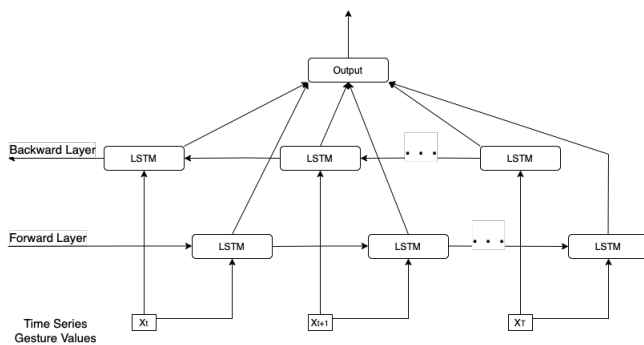


Fig. 4. BiLSTM architecture

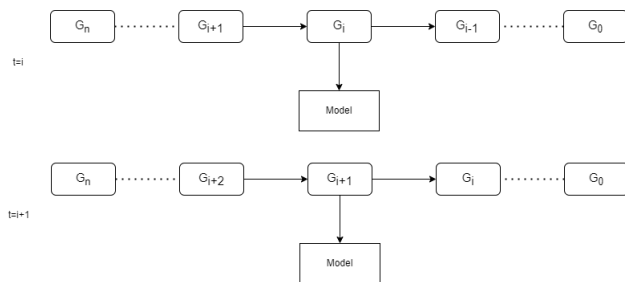


Fig. 5. Training process of BiLSTM model using online learning

IV. IMPLEMENTATION

Using the proposed model discussed in Section IV, a sensor-based HGR data glove was developed, is shown in Fig. 6 and its circuit diagram in Fig. 7, that is capable of capturing hand and finger movements.

The data glove was developed from scratch by making use of five 4.5” Flex Sensors, one for each finger and an MPU6050 GY-521 IMU sensor on the dorsal (back) side of the hand and an Arduino Mega 2560 microcontroller. The flex sensor is used to calculate the finger bend angle for each of the fingers and the IMU sensor is used to calculate the Euler Yaw, Pitch and Roll values. A second glove was developed in the same fashion to replicate the process and validate the feasibility of building of the glove. Both the gloves are shown in Fig. 6.

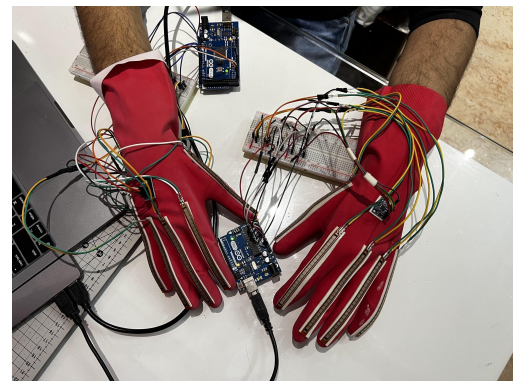


Fig. 6. Developed hand glove with embedded and sensing device to test the proposed model.

The raw data collected by the data glove is then processed by amplifying the signal and passed through low-pass filters to get rid of interferences and noises by using Moving Average Smoothing. It is then passed to the A/D converter of the microcontroller, capable of processing the time series data for machine learning algorithms to extract the features relating to the hand and finger movements carried out for the gestures intended by the users. This data is then processed through a neuro-computing system capable of classifying the gesture being performed in real time. The various hand gestures are mapped to their respective meanings and stored in a dictionary. These act as classes to create a labeled dataset for further training process.

Different gestures are considered for training the feed-forward neural network model integrated with DSTW and the BiLSTM model. For each gesture, 100 data frames were collected. After collecting the data, having taken 5 flex sensor angles shown in Fig. 8 captured at intervals of 20ms. A feed-forward neural network architecture is used with memory-based learning as the learning method to classify the gestures into their meanings. Dynamic spatio-temporal warping is used to compare time series data with spatial and temporal variations. The algorithm shows time complexity of  $O(m * n)$ . During the experimentation, numba python library for multi-processing was used to bring down the computation time from 1 hour to around 30 seconds. This resulted in increasing the

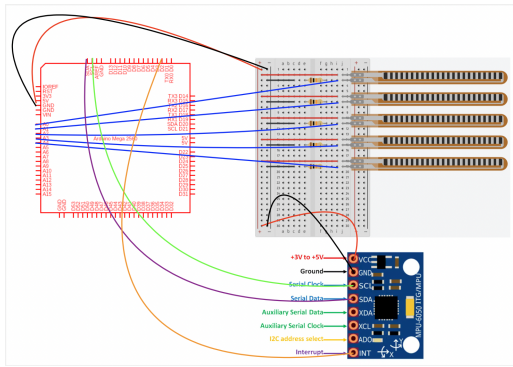


Fig. 7. Circuit diagram of the proposed HGR model.

TABLE I. GRAMMAR ERROR CORRECTION (GEC) OUTPUTS FOR SAMPLE INPUTS

Input	Restructured Sentence	GingerIt	Gramformer	Actual Sentence
Your phone number what	What your phone number	What's Your phone number	'What is your phone number?', 'What is Your phone number.'	'What is your number?'
I go theatre	I go theatre	I go theatre	'I go to the-atre.', 'I went to theatre'	I go to a the-atre

number of times the gestures were trained in a given period of time.

Once the gestures are classified, they act as the input to the NLP model that aims to determining the start and stop of a sentence and further segmentation is performed. They are then processed to correct the grammar and form meaningful sentences. DeepSegment [14], is a GloVe + BiLSTM Conditional Random Field (CRF) sequence model, to segment sentences with no punctuations. A trend was noticed in the ISL structure and grammar, where the interrogative sentences ended with the wh-pronoun, and the sentence was restructured by making use of the spaCy [15] library. However, the sentences were still incomplete, as there are no signs in ISL for articles. To combat this, grammar correcting libraries such as GingerIt [16], an LSTM based architecture, and Gramformer [17], a GPT-2 based model were used. Once this is done, the grammatically correct sentence then acts as an input to the text-to-speech generative algorithm which aims at conversion of the sentence to audio which can be played over a sound system / mobile device / speaker which helps the person at the receiving end understand what the user is trying to communicate. Table I illustrates the comparison between some of the grammar correction technologies like GingerIt and Gramformer. It was observed that Gramformer gave more accurate results with respect to grammar correction.

The data is fed into a BiLSTM model represented in Fig. 9 with 64 neurons in its input layer, which are BiLSTM in nature, another BiLSTM layer with 64 neurons as the hidden layer, and one neuron in the output layer, which gives us the class of the gesture. Adam Optimizer is used with Mean Square Error as the loss function.

The simulation of the collected Flex Sensor angles is shown

in Fig. 10. It represents the angles collected for 10 fingers on both hands using 10 graphs. To simulate the angle at which the finger is bent, two line segments are used having the same angle between them. One line segment is the line segment joining (0.0, 1.0) and (1.0, 1.0) and the other line segment is plotted using the Circle equation with center at (h, k),  $(x - h)^2 + (y - k)^2 = a^2$  for which the points on the circle are  $x = h + a \cos(\theta)$  and  $y = k + a \sin(\theta)$ . Therefore, for the circle with center at (1, 1) and radius  $a = 1$ , the circle equation becomes  $(x - 1)^2 + (y - 1)^2 = 1$  and for which the points on the circle becomes  $x = 1 + \cos(\theta)$  and  $y = 1 + \sin(\theta)$ . Since the angles being plotted are in the fourth quadrant,  $\theta = 360^\circ - \theta_f$  where  $\theta_f$  is the flex angle.

## V. RESULTS AND DISCUSSION

Performance of the model was carried for different gestures in Indian sign language involving hand and finger movements. Feed-forward neural network with memory-based learning integrated with Dynamic Spatio-temporal Warping and BiLSTM models were compared to classify the time series data collected from the data gloves. BiLSTM was found to have better performance during the experimentation. Transformer-based NLP model for grammar correction modules was found to work accurately for the contexts considered during the experimentation. Different performance parameters were considered in the experimentation, some of them are discussed below.

With the Feed forward neural network model integrated with DSTW, the data is split into train data - 70% and test data - 30% resulting in an accuracy of 96.74% as shown in Fig. 11.

With the BiLSTM model, the data is split into train data - 70% and test data - 30% resulting in an accuracy of 98.91% as shown in Fig. 12.

## VI. CONCLUSION

The development of wearable devices for sign language translation enhances the collaborative interaction with the augmentation of reality. In comparison to the vast number of existing technologies present to aid the deaf population, such specific wearable devices will be inexpensive and easy to use. Flex and IMU sensors calibrated and integrated by using hardware and software filters to collect accurate data of hand and finger movements with low amount of noise are proposed. Machine learning methodologies like Feed-forward neural network with memory-based learning and BiLSTM integrated with DSTW has enhanced the performance of gesture recognition as per the ISL video gestures. From the experimentation, BiLSTM model was found to have better performance with Feed-forward neural network with DSTW with accuracy of 98.91%. Multiprocessing features were used to improve the computation time during the classification. The gesture sequences with meanings used by NLP models for grammar correction were compared and it was found that transformer-based models outperformed other types of models and parsers. Obtained sentences were conveyed to the user through communicable media as per the user's requirement.

Gesture recognition devices can be used in multiple ways and in various fields of science, like in the medical field for

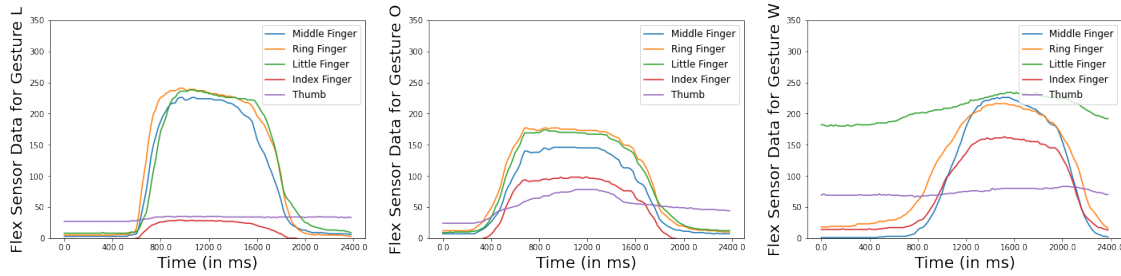


Fig. 8. Flex sensor data for gestures L, O and W.

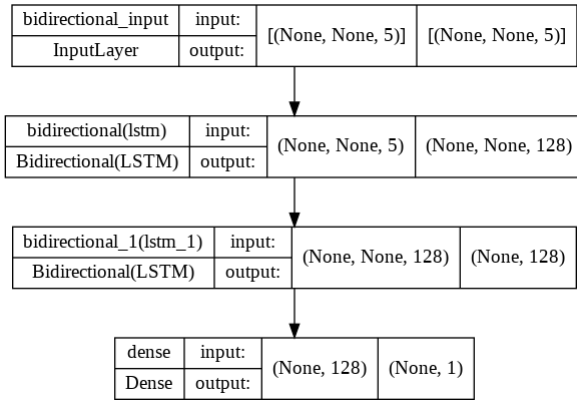


Fig. 9. BiLSTM model

	precision	recall	f1-score	support
Gesture L	0.9688	1.0000	0.9841	31
Gesture O	1.0000	0.9688	0.9841	32
Gesture W	1.0000	1.0000	1.0000	29
accuracy			0.9891	92
macro avg	0.9896	0.9896	0.9894	92
weighted avg	0.9895	0.9891	0.9891	92

Fig. 12. Classification report of the BiLSTM model

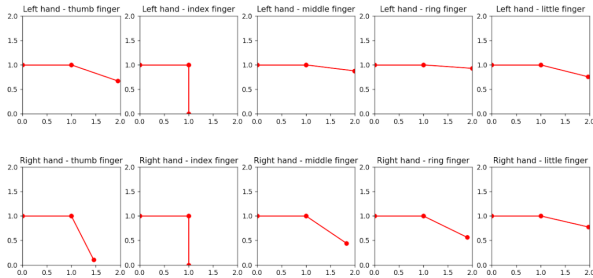


Fig. 10. Simulation of collected flex sensor data

	precision	recall	f1-score	support
Gesture L	1.0000	0.9706	0.9851	34
Gesture O	0.9655	0.9655	0.9655	29
Gesture W	0.9333	0.9655	0.9492	29
accuracy			0.9674	92
macro avg	0.9663	0.9672	0.9666	92
weighted avg	0.9681	0.9674	0.9676	92

Fig. 11. Classification report of the feed-forward neural network with DSTW model

remote operations/surgeries using HGR, as well as during the recovery of the patients either through physical rehabilitation or physical therapy that can be done remotely with gesture recognition. In other examples it can be used in interacting with a computer through HGR, virtual and augmented reality applications using gesture recognition, gesture controlled IoT home appliances and much more.

### A. Future Scope

This proposed model can be built on and adapted to enable the communication between varied disabled communities, like a person from the deaf community and a person from the blind community. The placement of the sensors on the glove could be optimized to ensure that the data collected is representative of the hand gesture being performed. To make the glove more responsive and usable in real-world scenarios, the data processing and classification can be improved to work in real-time. The number of classes of gestures that the glove is able to recognize should include all the gestures in the ISL.

### ACKNOWLEDGMENT

The authors thank the Artificial Intelligence and Robotics Technology Park (ARTPARK) division of the Indian Institute of Science (IISc), Bangalore for providing resources and technical support. We also acknowledge the Information Science and Engineering department of M S Ramaiah Institute of Technology for their support.

### REFERENCES

[1] United Nations, International day of sign languages, <https://www.un.org/en/observances/sign-languages-day, 2022>

- [2] Faculty of Disability Management and Special Education, "Indian Sign Language Portal", [https:// indiansignlanguage.org/](https://indiansignlanguage.org/), 2022
- [3] Chuang, W.C., Hwang, W.J., Tai, T.M., Huang, D.R., Jhang, Y.J. Continuous finger gesture recognition based on flex sensors, 2019
- [4] Zhou, Z., Chen, K., Li, X., Zhang, S., Wu, Y., Zhou, Y., Meng, K., Sun, C., He, Q., Fan, W., Fan, E., Lin, Z., Tan, X., Deng, W., Yang, J., Chen, J. Sign-to-speech translation using machine-learning-assisted stretchable sensor array, 2020
- [5] Tai, T.M., Jhang, Y.J., Liao, Z.W., Teng, K.C., Hwang, W.J. Sensor-based continuous hand gesture recognition by long short-term memory, 2018
- [6] Zheng, J., Zhao, Z., Chen, M., Chen, J., Wu, C., Chen, Y., Shi, X., Tong, Y. An improved sign language translation model with explainable adaptations for processing long sign sentences, 2020
- [7] Lang, B. Parsing incomplete sentences, 1988
- [8] Alikaniotis, Dimitrios & Raheja, Vipul. The Unreasonable Effectiveness of Transformer Language Models in Grammatical Error Correction, 2019
- [9] Patil, S., Bidari, I., Sunag, B., Gulahosur, S.V., Shettar, P. Application of hmi technology in automotive sector, 2016
- [10] Government of India, Indian Sign Language Research and Training Center (ISLRTC), [http://www.islrte.nic.in./](http://www.islrte.nic.in/), 2022
- [11] Jangyodsuk, P., Conly, C., Athitsos, V. Sign language recognition using dynamic time warping and hand shape distance based on histogram of oriented gradient feature, 2014
- [12] Barth, J., Oberndorfer, C., Pasluosta, C., Schülein, S., Gassner, H., Reinfelder, S., Kugler, P., Schuldhuis, D., Winkler, J., Klucken, J., Eskofier, B.M. Stride segmentation during free walk movements using multi-dimensional subsequence dynamic time warping on inertial sensor data, 2015
- [13] Hameed, Z., Garcia-Zapirain, B. Sentiment classification using a single-layered bilstm model, 2020
- [14] notAI tech, "Deepsegment", <https://github.com/notAI-tech/deepsegment>, 2020
- [15] Explosion, "spacy", <https://github.com/explosion/spaCy>, 2022
- [16] Azd325, "Gingerit" <https://github.com/Azd325/gingerit>, 2022
- [17] Damodaran, P. "Gramformer", <https://github.com/PrithivirajDamodaran/>, 2022

# Delivery Management System based on Blockchain, Smart Contracts and NFT: A Case Study in Vietnam

Khiem Huynh Gia<sup>1</sup>, Luong Hoang Huong<sup>2</sup>, Hong Khanh Vo<sup>3</sup>, Phuc Nguyen Trong<sup>4</sup>, Khoa Tran Dang<sup>5</sup>,  
Hieu Le Van<sup>6</sup>, Loc Van Cao Phu<sup>7</sup>, Duy Nguyen Truong Quoc<sup>8</sup>, Nguyen Huyen Tran<sup>9</sup>, Anh Nguyen The<sup>10</sup>,  
Huynh Trong Nghia<sup>11</sup>, Bang Le Khanh<sup>12</sup>, Kiet Le Tuan<sup>13</sup>, Nguyen Thi Kim Ngan<sup>14</sup>  
FPT University, Can Tho City, Viet Nam<sup>1-13</sup>  
FPT Polytechnic, Can Tho City, Viet Nam<sup>14</sup>

**Abstract**— Current traditional shipping models are increasingly revealing many shortcomings and affecting the interests of sellers and buyers due to having to depend on trusted third parties. For example, the Cash-on-Delivery (CoD) model must depend on the carrier/shipper, or the Letter-of-Credit (LoC) model depends on the place of the Letter certification (i.e., bank). There have been many examples demonstrating the riskiness of the two models above. Specifically, in developing countries (e.g., Vietnam), the demand for exporting goods and trading between sellers and buyers have not yet applied the benefits of current technology to improve traditional shipping models. Two typical examples in the last five years that have demonstrated the risks of both sellers and buyers when applying CoD and LoC models are the problem of keeping the money of the seller of GNN Expresses (2017) as well as risks in losing control of 4 containers of cashew nuts when exporting from Vietnam to Italy (2021). A series of studies have proposed solutions based on distributed storage, blockchain, and smart contracts to solve the above problems. However, the role of the shipper has not been considered in some approaches or is not suitable for deployment in a developed country (i.e., Vietnam). In this paper, we propose a combination model between the traditional CoD model and blockchain technology, smart contracts, and NFT to solve the above problems. Specifically, our contribution includes four aspects: a) proposing a shipping model based on blockchain technology and smart contracts; b) proposing a model for storing package information based on Ethereum's NFT technology (i.e. ERC721); c) implementing the proposed model by designing smart contracts that support the creation and transfer of NFTs between sellers and buyers; d) deploy smart contracts on four EVM-enabled platforms including BNB Smart chain, Fantom, Celo, and Polygon to find a suitable platform for the proposed model.

**Keywords**—Letter-of-Credit; cash-on-delivery; blockchain; smart contract; NFT; Ethereum; Fantom; Polygon; Binance Smart Chain

## I. INTRODUCTION

Today, delivery from sellers and buyers is growing constantly where many traditional shipping models (e.g. Cash-on-delivery (CoD)[1], Letter-of-Credit (LoC) [2]) to transportation models applying advanced technologies (e.g. blockchain, RFID, smart contracts) [3]. The main purpose of these models is to shorten delivery and pickup times and make it possible for sellers and buyers to trace the location and expected time of delivery [4]. For developing countries (i.e. Vietnam), the current delivery and receipt process still applies the traditional delivery and receipt model. Specifically, with the steps of delivery and receipt of goods within a city or between cities,

the common model applied between sellers and buyers is CoD - the buyer will send money directly to the delivery party. The seller is responsible for shipping the item to the buyer through the shipping company (i.e. shippers). For cross-border transactions (i.e. between Vietnam and other countries - in the region or in the world), the commonly used model is LoC. All exchanges between the two parties are recorded and authenticated by a trusted third party (i.e. a bank).

However, both models have a lot of risks. With the LoC model, in case either party (i.e. exporter and importer) loses the original document (i.e. letter), the possibility of a loss of goods/money is very high. Because the seller only receives the full amount when the buyer receives the goods. For the CoD model, the seller must accept the risk of trusting and authorizing the shipping company (i.e. money, goods). Specifically, the number of goods that will be returned to the seller by the shipping company after a fixed period of time (e.g. monthly, quarterly) or a certain amount is reached. For this type of shipping, the risk for the seller is very high because the shipping companies can use their money for something other than sending the money back to the seller. Most of today's processing is based on an agreement between two parties (i.e. seller and carrier).

The application of traditional freight forwarding and payment models in Vietnam is facing a number of problems in the past five years. For the LoC model, in case one of the parties loses the original documents (i.e. letter), the possibility of losing goods/money is very high. A specific example happened in 2021 for the cashew nut export model from Vietnam to Italy. Specifically, four out of 100 containers of cashews exported from Vietnam to Italy are at risk of being lost because exporting companies in Vietnam cannot present original documents.<sup>1</sup> Fortunately this issue was resolved with the involvement of the State Department Vietnam and the Consulate General of Vietnam in Italy. For the traditional CoD model, an example of a seller being stolen by GNN Express occurred in Vietnam in 2017 and 2018.<sup>2</sup> Specifically, the entire amount of goods worth about \$154,900 was not transferred to the seller but was used by GNN Express for other purposes.

To solve the above problem, a series of models have proposed Blockchain technology and smart contracts to easily trace the origin of packages as well as reduce the risk from

<sup>1</sup><https://vietnamnet.vn/en/100-containers-of-cashew-nuts-exported-to-italy-suspected-of-being-scammed-821553.html>

<sup>2</sup><https://vir.com.vn/gnn-scandal-rocks-delivery-segment-62710.html>

untrusted parties. This model (i.e. in theory) can operate independently of a third party (i.e. no trusted third party required). However, if there is a conflict between the seller and the buyer, the arbitration must still be requested (i.e. the arbitrator receives a fee known as the conciliation fee). One of the popular examples of this pattern is Bitcoin[5], and Ethereum [6]. Specifically, all transactions will be routed through the seller/buyer's wallet address [7]. In addition, state-of-the-art has extended the above two models to solve the traditional freight problem (see the Related work section for more details). For research directions related to the transportation of goods for developing countries (i.e. Vietnam). There are not many in-depth approaches to ensure the transportation of traditional goods (i.e. see related work for more details).

In addition to the problems related to the implementation of a blockchain-based shipping system and smart contracts in Vietnam, the above method encounters some limitations when it ignores an important group of people who are shippers. Specifically, the role of this user group is ignored when shipping goods from the seller and the buyer. This can affect the arbitration process when there is any conflict between the seller, and buyer [8]. This has prompted several theoretical studies that combine the role of shippers using blockchain technology and smart contracts [9].

However, the above approaches still face a challenge when transporting packages between different shippers or shipping companies in Vietnam. In particular, packages can get damaged in transit - it's difficult to know who is responsible (i.e. seller, buyer, or carrier) as all interactions are done on paper and are not legally binding. Therefore, in this study, we aim to determine the package contents and related information when moving from seller to buyer. Specifically, we apply NFT technology to generate package-related information (i.e. sender, recipient, order content, weight, estimated delivery time, etc.). When the buyer receives the package, he can check the information sent from the seller. This technology is easy to deploy in developing countries because it does not require too high of a technology-based infrastructure to maintain the system. Some countries have implemented delivery models based on Blockchain, smart contracts, and NFT technologies.

Therefore, our contribution includes four aspects: a) proposing a shipping model based on blockchain technology and smart contracts applicable to the delivery environment in Vietnam; b) proposing a model for storing package information based on Ethereum's NFT technology (i.e. ERC721); c) implementing the proposed model by designing smart contracts that support the creation and transfer of NFTs between sellers and buyers; d) deploying smart contracts on four EVM-enabled platforms including BNB Smart chain, Fantom, Celo, and Polygon to find a suitable platform for the proposed model.<sup>3</sup>

The rest of the paper consists of seven parts. After the introduction is the related work section, which presents state-of-the-art with the same research problem. The next two sections present our approach and the proposed model implementation (i.e., Sections III, IV). To demonstrate our effectiveness, Section V presents our evaluation steps in different scenarios

<sup>3</sup>We do not implement smart contracts on ETH because the execution fee of smart contracts is too high.

before making comments in Section VI. The Section VII summarizes and outlines the next steps for development.

## II. RELATED WORK

This section presents approaches from traditional shipping (i.e. post office, third party - courier companies, and e-commerce platforms) to proposed theoretical models based on technology. Blockchain and smart contracts.

### A. Traditional Delivery Method

The demand for transporting goods is increasingly diverse (especially in developing countries). In this paper, we summarize the common approaches applied to Vietnam. In particular, the simplest approach is based on the postal system, which plays an important role in transporting goods. All information about the sender and receiver is done through a third party (i.e. bank employee)<sup>4</sup>. In the traditional model, shipping and delivery management is managed in a decentralized and centralized manner. Specifically, each city and province has a post office that acts as a receiving and transit point to sub-post offices in the districts (i.e. low-level). The post office in the district receives and sends the sender's item to a corresponding location on the package.

The benefits of this approach are low transportation costs and no need for smart devices/systems (i.e. smartphones, sensors). However, it also brings a lot of disadvantages i) it is difficult for users to get information about their packages; ii) long transit times; iii) if the recipient cannot be found, it will take time to send it back to the sender; iv) requires a large number of service workers due to lack of technology support (these risks are introduced in detail in the Introduction section).

In a more advanced approach, courier companies were born to ensure on-time delivery - with a higher fee. Specifically, courier companies will ensure delivery in the shortest time. A few popular courier companies in Vietnam include two groups: domestic courier companies: Viettel Post, Vietnam Post J&T Express<sup>5</sup> and has its headquarters abroad: Fedex<sup>6</sup>, ASL<sup>7</sup>, DHL<sup>8</sup>.

In addition, e-commerce platforms have also changed the shipping trend in Vietnam in the past 10 years [10]. Specifically, e-commerce platforms act as intermediaries to link sellers and buyers. In other words, they bridge the gap between the two - finding products required by buyers and finding potential user groups for sellers. For sellers, they put their products on e-commerce platforms instead of having to open a store in real life. Buyers also save time shopping and commuting. Famous e-commerce platforms in Vietnam today include Lazada<sup>9</sup>, Tiki<sup>10</sup>, SenDo<sup>11</sup>, Amazon<sup>12</sup>, and Shopee<sup>13</sup>. In

<sup>4</sup><http://www.vnpost.vn/en-us/dich-vu/chi-tiet?id/183/key/parcel-post-service>

<sup>5</sup><https://viettelpost.com.vn/activity-news/top-5-most-prestigious-and-quality-delivery-services-in-vietnam/>

<sup>6</sup><https://www.fedex.com/en-vn/home.html>

<sup>7</sup><https://www.asl-corp.com.vn/>

<sup>8</sup><http://dhlexpress.vn/>

<sup>9</sup><https://www.lazada.vn/>

<sup>10</sup><https://tiki.vn/>

<sup>11</sup><https://www.sendo.vn/>

<sup>12</sup><https://amazon.vn/>

<sup>13</sup><https://shopee.vn/>



addition, the role of freight is guaranteed by shipping companies of a small size (i.e. between cities or within the same city). These people will be the ones who directly receive the money from the buyer and deliver it back to the seller (i.e. Cash-on-delivery). The disadvantage of this model is that the seller can lose money when the delivery company goes bankrupt. The closest example to this is GNN Express<sup>14</sup>. Specifically, the company GNN Express took money from the seller to make up for the company's loss. Currently, this company is bankrupt and unable to pay. Another example related to cross-border freight took place in early 2021 between Vietnam and Italy. An Italian company exploited vulnerabilities in the Letter-of-Credit delivery method to hijack four containers of cashews. While the above issues have been addressed, we clearly see undeniable flaws in traditional delivery models.

### B. Delivery Method based on Blockchain Technology

Previous approaches to deploying their proof-of-concept are mainly on the two platforms Hyperledger Fabric and Ethereum

1) *Ethereum*: The common point of these protocols is the use of ETH as the mainstream payment currency. For example, the Ethereum ecosystem has proposed localEthereum a method to facilitate transactions or DeFi Dapps between providers and claimants [7]. In addition, a new protocol that aims to help ship products from supplier to requester [11] has mined an ETH-based transaction to propose a COD/LOC mechanism. Similarly, OpenBazaar [12] is developed based on the extension of the Ethereum ecosystem, supporting sellers and buyers in the process of exchanging goods. Specifically, the transaction is identified based on the wallet address (i.e. transfers funds from the buyer's wallet address to the seller's wallet address) and is easily authenticated by the relevant parties without the need for a trusted third party. However, unlike localEthereum, OpenBazaar involves three parties: the provider, the requestor, and the moderator (that is, a new control role). Specifically, instead of just applying the entire transaction based on a default smart contract, OpenBazaar supports middleman (i.e. owner of smart contracts) to define corresponding policies and penalties to control the process. the process of transporting goods to determine the subject of compensation if an incident occurs. However, these systems suffer from a single point of failure problem - ignoring the role of the carrier. Specifically, the carrier will not be affected if there is any conflict between the seller and the buyer (i.e. purchase of poor quality goods, loss of the package) - even if it is the fault of the carrier's side [13].

2) *Hyperledger Fabric*: The common point between the approaches developed on the Hyperledger Fabric platform is to highlight the role of the carrier - the carrier is also responsible for the exchange of goods. Specifically, policies designed in the chain code (i.e. smart contracts) identify violations by all three groups of participants including seller, buyer, and person/shipping company[14]. For example, Son et al. [15] have proposed a baseline model of carriers, sellers, and buyers to replace the current CoD model. However, this approach requires no deposit constraints - conflicts are handled by a trusted third party (i.e. arbitration). To solve this problem, a

few (e.g. [16], [17]) have proposed a model that combines blockchain technology and smart contracts in shipping packages between sellers and buyers. In which, the violations of the parties involved are sanctioned with a fee deducted from the previous deposit. Specifically, [17] supports a variety of shippers - multiple shippers can be involved in the delivery instead of just one object/shipping company as [16] suggests. To accommodate transit times (i.e. not limited by geographical distance), Duong et al. [18] proposes a new approach based on multi-section. For the purpose of supporting payment on a variety of platforms - users can create an account at one e-commerce platform and make purchases at all other exchanges. In addition to solutions to support sellers and buyers in the process of exchanging goods, Son et al. [1] also proposes a cross-platform model connecting many different markets, where sellers and buyers are not limited by payment tools between e-commerce platforms. Ha et al. [19] has proposed a personal information protection model (i.e. seller and buyer) based on the access control model [20], [21]. Specifically, sensitive information such as addresses and phone numbers can only be accessed by those with [22] permissions. This solution greatly helps in protecting the personal information of the seller and the buyer because this information is always available on other electronic exchanges (even those that do not play a role in the corresponding transaction). For cross-border transactions, Khoi et al. [2] have proposed a mechanism for transporting goods between countries that are not dependent on a trusted third party (i.e. bank) called Letter-of-Credit. This approach strives for an open policy across countries (i.e. unconstrained by the policies of one country, geographic region [23], [24]).

However, all of the above approaches are still bound when transferring orders between shippers or between seller - shipper or shipper - buyer. In particular, the package may be damaged in transit and it is difficult to determine who is responsible for this damage. To solve this problem, we propose a shipping model based on a combination of some outstanding technologies today, such as blockchain, smart contracts, and NFT. The next section describes in detail the model and how to operate the system.

## III. LETTER-OF-CREDIT CHAIN ARCHITECTURE

### A. The Traditional Model of Freight Transport

In this section, we analyze the popular traditional freight model in Vietnam (i.e. CoD). Fig. 1 shows the six steps of shipping goods from the seller and the buyer. Specifically, the seller and the buyer discuss the price and form of payment. In Vietnam, this process is done based on social networking platforms (i.e. Facebook) or based on e-commerce platforms (i.e. sellers share product information and prices - buyers - buyers) product selection). The seller checks whether the product is in stock (step 2) before packing the product (step 3). In step 4, the seller chooses a reputable shipping company because the shipping company holds the customer's payment before returning it to them. Step 5 shows the process of shippers coming to pick up the goods from the seller and get the shipping address. Finally, the shippers ship the item to the buyer (step 6). The difficulties and risks of the traditional CoD model have been analyzed by us in the Introduction and Related work sections.

<sup>14</sup><https://vir.com.vn/gnn-scandal-rocks-delivery-segment-62710.html>

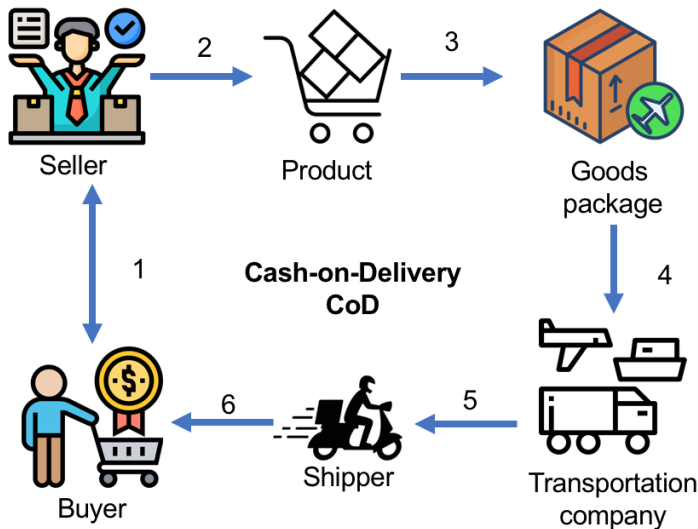


Fig. 1. The Traditional Model w.r.t cash-on-delivery of Freight Transport.

For the process of transporting goods between countries, we have an article that specifically analyzes the risks of the traditional model (i.e. LoC) and presents the blockchain-based approach and smart contracts in the process. [2]. In this section we focus on solving the problem of transporting goods within the territory of Vietnam (i.e. between cities or within the same city). Depending on the shipping distance, step 6 (i.e. Fig. 1) includes one or more shippers. The next section presents our proposal process based on a combination of Blockchain technology, smart contracts, and NFT.

#### B. Cargo Transport Model based on Blockchain Technology, Smart Contracts and NFT

Fig. 2 presents our proposed model based on blockchain technology, smart contracts, and NFT (i.e. consisting of nine steps). We assume that the seller and the buyer have agreed on the delivery method as well as the price for the product. In this section, we design the system to move goods from seller to buyer. Specifically, step 1 shows the seller to pack the product and prepare the necessary declaration information related to the package (i.e. weight, unit price, item type, etc. see Implement section) in step 2 ( i.e. based on system support services). These services call the respective functions in smart contracts to create a shipping contract (step 3). The seller selects the corresponding shipping company (step 4). Step 5 presents the deposit confirmation process for each participating group. In it, the seller still has to enter an amount of money to ensure delivery of the correct goods and of the right quality (i.e. if the seller intentionally violates the information provided in smart contracts - step 2 - this amount is used to compensation to the carrier and the buyer). For shippers, their management company must pay a guarantee fee to avoid the shipper losing the goods or in case the company goes bankrupt before the time to refund the buyer (i.e. the amount of money). Their deposit depends on the exchange between the seller and the shipping company). In the event of a conflict (i.e. shippers lose or damage goods or the shipping company goes bankrupt) smart contracts automatically transfer the company's deposit to the seller via their address. The

buyer's deposit includes the shipping fee and part of the deposit of the product (i.e. depending on the agreement between the seller and the buyer). In case the buyer refuses to purchase the product, the shipping fee and deposit of the product are automatically transferred to the address of the shipping company and the seller, respectively. The above conventions are monitored through protocols designed on smart contracts and stored on NFTs with the consent of all three parties (i.e. seller, buyer, and shipping company) - step 6. Step 7 presents the shipping process between shippers (i.e. depending on the distance the number of shippers is 1 person or many). This process is managed and operated by shipping companies. Step 8 presents the final shipment (i.e. the buyer receives the item) - confirmation from the buyer that the package is correct and that the information is stored in the NFT. The transactions are conducted specifically, the buyer pays the remaining amount of the product; the shipping company receives the deposit and shipping; The seller receives the money for the sale of the product. Risks and breaches of contract are resolved based on the cases designed in step 5. Finally, smart contracts update transactions to the distributed ledger and prepare for a new shipping process.

#### IV. IMPLEMENTATION

Our reality model focuses on two main purposes i) data manipulation (i.e. package) - initialization, query and update - on blockchain platform and ii) generation of NFT for each order for easy traceability by sellers and buyers (i.e. product reviews before and after delivery).

##### A. Initialize Data / NFT

Fig. 3 shows the data row startup step packages. These package types include information related to the sender (i.e. receiving address, weight, type of item), and the recipient (i.e. receiving address, expected delivery time). In addition, the exchange and receipt of goods require an account deposit of all three parties depending on the purpose and transaction between the parties to ensure automatic conflict resolution on smart contracts. . In addition, information about which carrier belongs to which company, time, and place of delivery and collection is also added to the package's metadata. This is extremely important in cases where more than one shipper is involved in the transportation of household goods (i.e. the same or different shipping company). For storage, services support concurrent storage (i.e. processing partitions as a peer-to-peer network) on a distributed ledger - Supports more than one user for concurrent storage, the speed drop of the whole system. In general, the package data is organized as follows:<sup>15</sup>

```
goodsObject = {  
  "goodsID": goodsID,  
  "deliveryCompanyID": deliveryCompanyID,  
  "shipperID": shipperID,  
  "type": type of goods,  
  "buyerID": buyerID,  
  "sellerID": sellerID,  
  "quantity": quantity,  
}
```

<sup>15</sup>The information related to the system participants is not listed in the article. Readers can read more about the group's previous research at [25], [18], [1].

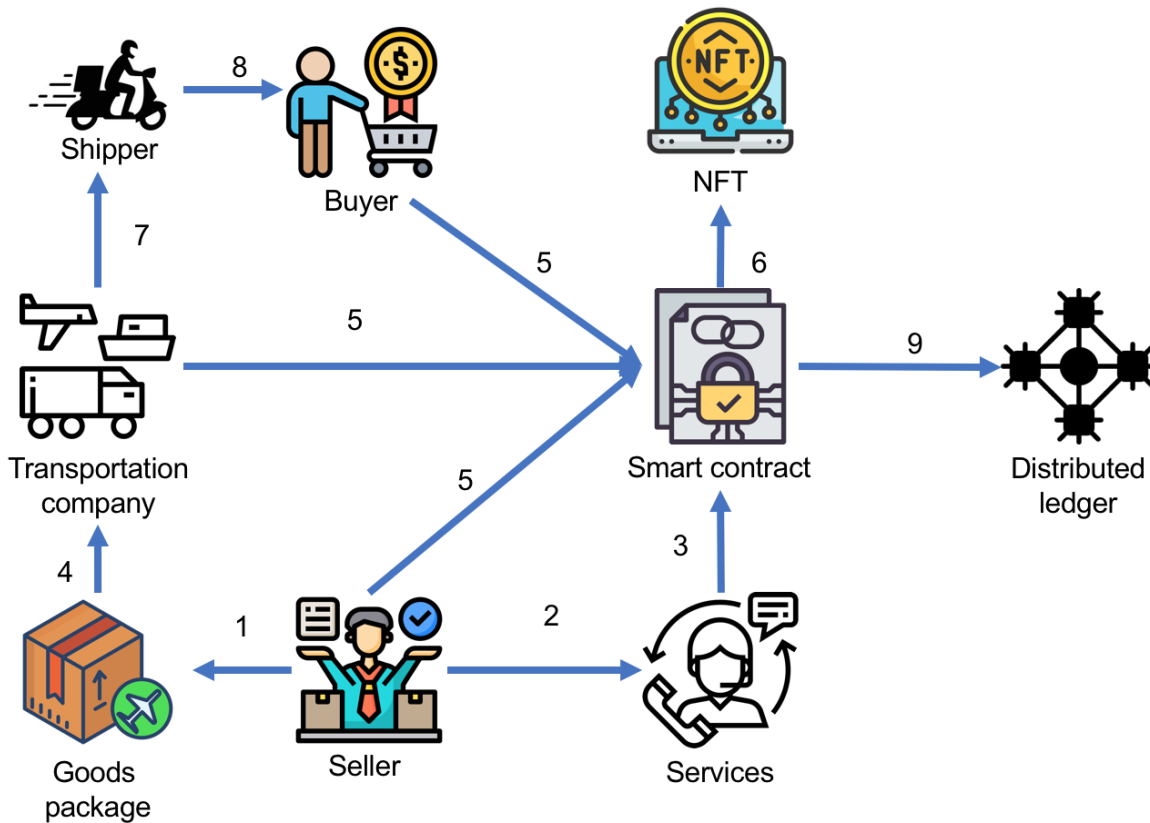


Fig. 2. Cargo Transport Model based on Blockchain Technology, Smart Contracts and NFT.

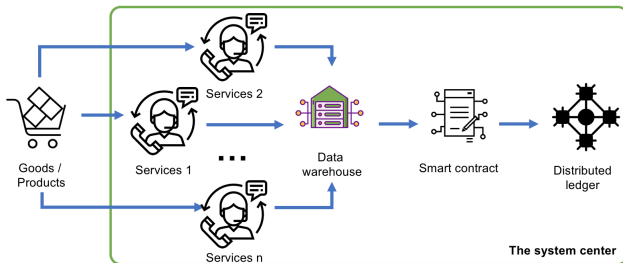


Fig. 3. Initialize Data/NFT.

of orders (e.g. 10) as well as which “packageID” they are assigned to. After receiving packages from the seller, the shipper checks them for compliance and waits for validation before syncing up the chain (i.e., temporarily stored on the data warehouse). Then the pre-designed constraints in Smart Contracts are called through the API (i.e. name of the function) to sync them up the chain. This inspection role is extremely important because they directly affect the shipping process of goods, as well as the premise for conflict resolution when any problems arise (e.g. damage. goods, lost packages). For the processes of initiating NFTs for orders (i.e. determining the deposit amount of the seller, buyer, and carrier), the content of the NFT is defined as follows:

```
"unit": unit,
"packageID": packageID,
"addressReceived": received address,
"addressDelivery": delivery address,
"time": estimated delivery time,
"location": location,
"state": Null
};
```

Specifically, in addition to information for content extraction (i.e. place of origin, weight, item type, etc.), we also store information regarding the status of the package at “addressReceived” (i.e. “state” - default value is Null). Specifically, “state” changes to 1 if the corresponding package has been received and shipped by the shipping company (i.e. “shipperID”); value 0 - pending (i.e. waiting for the shipper to pick up the item). In addition, “unit” stores the number

```
NFT PACKAGE = {
"shipperID": shipperID,
"sellerID": sellerID,
"buyerID": buyerID,
"packageID": packageID,
"type": type of goods,
"quantity": quantity,
"addressReceived": received address,
"addressDelivery": delivery address,
"depositShipper": deposit of shipper,
"depositSeller": deposit of seller,
"depositBuyer": deposit of buyer,
"time": estimated delivery time
};
```

The information on the NFT contributes to conflict reso-

lution (e.g. delivery delays). Definitions related to stakeholder deposit have been defined in our previous articles.

B. Data Query

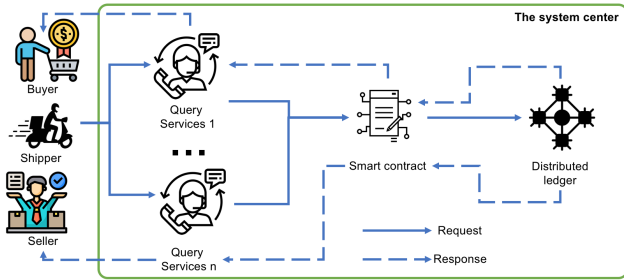


Fig. 4. Data Query.

Similar to the data initialization steps, the data query process also supports simultaneous access by many participants to the system (i.e. distributed model). Support services receive requests from shippers or sellers/buyers to access data (i.e. respective packages). Depending on the query object, we have different access purposes. Specifically, shippers query for the purpose of identifying consignee information and their addresses. In contrast, sellers/buyers view the status of their orders (i.e. after being delivered and received) as well as handling conflicts when something goes wrong. Fig. 4 shows the steps to query the order data. These requests are sent as requests (i.e. pre-designed services as API calls) from users to smart contracts available in the system (i.e. name of the function) before retrieving data from the distributed ledger. All retrieval requests are also saved as query history for each individual or organization. For a shipping process that involves multiple discounts (i.e. multiple shippers delivering and receiving the goods before reaching the buyer's address), NFTs are similarly created between shippers (i.e. within or different shipping companies). In case the corresponding information is not found (e.g. wrong ID), the system will send a message not found results. For the NFT query process, all support services are provided as APIs.

C. Data Updated

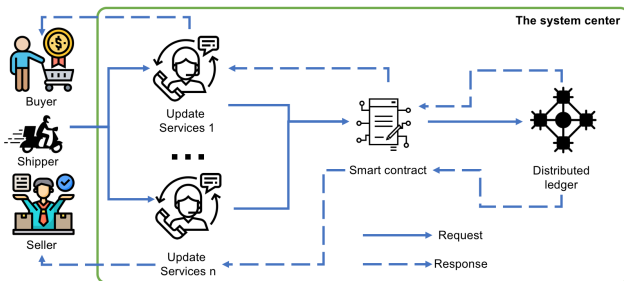


Fig. 5. Data Updated.

The data update procedure is invoked only after verifying that the data exists on the chain (i.e. after executing the corresponding data query procedure). In this section, we assume that the search data exists on the string. Where none exists, the system sends the same message to the user (see

IV-B for details). Similar to the two processes of query and data initialization, we support update services in the form of APIs to receive requests from users before passing them to smart contracts (i.e., name of the function) for processing. The purpose of this process is to update the status of the package during transit as well as handle conflicts when something goes wrong (i.e. a combination of smart contracts and NFT). Fig. 5 shows the process of updating order data. For NFTs (i.e., available), the update process includes only the transfer from the owner's address to the new address (i.e., new owner). If any information is updated on an existing NFT, it will be stored as a new NFT (see IV-A for details).

V. EVALUATION

Transaction Hash	Method	Block	Age	From	To	Value	[Txn Fee]
0x44a035eb56eead581...	Transfer	24868241	1 day 18 hrs ago	0xcaa9c5b45206e083f4f...	0x2ec70233d91ade867...	0 BNB	0.0027039
0x3c115b0c294ec94503...	Mint	24868230	1 day 18 hrs ago	0xcaa9c5b45206e083f4f...	0x2ec70233d91ade867...	0 BNB	0.00109162
0x03154e0340b96c4c37...		24868219	1 day 18 hrs ago	0xcaa9c5b45206e083f4f...		0 BNB	0.0273184

Fig. 6. The Transaction info (e.g., BNB Smart Chain).

More Info

My Name Tag: Not Available

Contract Creator: 0xcaa9c5b45206e083f4f... at txn 0x03154e0340b96c4c37...

Token Tracker: NFT PACKAGE (PACKAGE)

Fig. 7. NFT Creation.

Transaction Hash	Age	From	To	Token ID	Token
0x44a035eb56eead581...	1 day 18 hrs ago	0x2ec70233d91ade867...	OUT: 0xcaa9c5b45206e083f4f...	1	ERC-721: NFT..._AGE
0x3c115b0c294ec94503...	1 day 18 hrs ago	0x00000000000000000000...	IN: 0x2ec70233d91ade867...	1	ERC-721: NFT..._AGE

Fig. 8. NFT Transfer.

Because of the connection model between seller - shipper - buyer and support for payment currency (i.e. consensus protocol), we implement the proposed model on blockchain platforms that support EVM instead of mining platforms. belongs to the Hyperledger eco-system. In addition, assessments based on system responsiveness (i.e. number of requests responded successfully/failed, system latency - min, max, average) were evaluated by us in the tests. previous research paper. Therefore, in this paper, we determine the suitable platform for our proposed model. Specifically, we install a recommendation system on four popular blockchain platforms today, supporting Ethereum Virtual Machine (EVM), including Binance Smart Chain (BNB Smart Chain)<sup>16</sup>; Polygon<sup>17</sup>; Fantom<sup>18</sup>; and Celo<sup>19</sup>. Our implementations on these four platforms are also shared

<sup>16</sup><https://github.com/bnb-chain/whitepaper/blob/master/WHITEPAPER.md>  
<sup>17</sup><https://polygon.technology/lightpaper-polygon.pdf>  
<sup>18</sup><https://whitepaper.io/document/438/fantom-whitepaper>  
<sup>19</sup><https://celo.org/papers/whitepaper>

TABLE I. TRANSACTION FEE

	Contracts Creation	Create NFT	Transfer NFT
BNB Smart Chain	0.02731184 BNB (\$8.32)	0.00109162 BNB (\$0.33)	0.00057003 BNB (\$0.17)
Fantom	0.009576994 FTM (\$0.001850)	0.000405167 FTM (\$0.000078)	0.0002380105 FTM (\$0.000046)
Polygon	0.006840710032835408 MATIC(\$0.01)	0.000289405001852192 MATIC(\$0.00)	0.000170007501088048 MATIC(\$0.00)
Celo	0.0070974384 CELO (\$0.004 )	0.0002840812 CELO (\$0.000 )	0.0001554878 CELO (\$0.000 )

TABLE II. GAS LIMIT

	Contracts Creation	Create NFT	Transfer NFT
BNB Smart Chain	2,731,184	109,162	3,000,000
Fantom	2,736,284	115,762	72,803
Polygon	2,736,284	115,762	72,803
Celo	3,548,719	142,040	85,673

TABLE III. GAS USED BY TRANSACTION

	Contracts Creation	Create NFT	Transfer NFT
BNB Smart Chain	2,731,184 (100%)	109,162 (100%)	57,003 (1.9%)
Fantom	2,736,284 (100%)	115,762 (100%)	68,003 (93.41%)
Polygon	2,736,284 (100%)	115,762 (100%)	68,003 (93.41%)
Celo	2,729,784 (76.92%)	109,262 (76.92%)	59,803 (69.8%)

as a contribution to the article to collect transaction fees corresponding to the four platforms' supporting coins<sup>20</sup>, i.e. BNB<sup>21</sup>; MATIC<sup>22</sup>; FTM<sup>23</sup>; and CELO<sup>24</sup>. For example, Figure 6 details our three evaluations of a successful installation on BNB Smart Chain (i.e. similar settings are shown for the other three platforms). Our implementations to evaluate the execution cost of smart contracts (i.e. designed based on Solidity language) run on testnet environments of four platforms in order to choose the most cost-effective platform to deploy. reality. Our detailed assessments focus on the cost of performing contract creation, NFT generation (see Fig. 7) and NFT retrieval/transfer (i.e. NFT ownership update - see Fig. 8) presented in the respective subsections related to i) Transaction Fee; ii) Gas limit; iii) Gas Used by Transaction; and iv) Gas Price.

#### A. Transaction Fee

Table I shows the cost of creating contracts for the four platforms. It is easy to see that the highest transaction fee of the three requirements is contract creation for all four platforms. In which, the cost of BNB Smart Chain is the highest with the highest cost when creating a contract is 0.02731184 BNB (\$8.32); whereas, the lowest cost recorded by the Fantom platform with the highest cost for contract initiation is less than 0.009576994 FTM (\$0.001850). Meanwhile, the cost to enforce Celo's contract initiation requirement is lower than Polygon's with only \$0.004 compared to \$0.01. For the remaining two requirements (Create NFT and Transfer

NFT), we note that the cost of implementing them for all three platforms, Polygon, Celo, and Fantom is very low (i.e. negligible) given the cost. trades close to \$0.00. However, this cost is still very high when deployed on BNB Smart Chain with 0.00109162 BNB (\$0.33) and 0.00057003 BNB (\$0.17) for Create NFT and Transfer NFT, respectively.

#### B. Gas Limit

Table II shows the gas limit for each transaction. Our observations show that the gas limits of the three platforms (i.e. BNB, Polygon, and Fantom) are roughly equivalent - where Polygon and Fantom are similar in the first two transactions. Particularly in the third transaction, BNB's gas limit was extremely high at 3,000,000. While the gas volume of Polygon and Fantom is equivalent to 72,803. The other platform (i.e., Celo) has the highest gas limit in the first two transactions with 3,548,719; 142,040, respectively.

#### C. Gas Used by Transaction

Table III shows the amount of gas used when executing the transaction (i.e. what percentage of gas in total gas is shown in Table II). Specifically, the three platforms BNB, Polygon, and Fantom use 100% of the Gas Limit for the two Contracts Creation and Create NFT transactions. Meanwhile, Celo uses 76.92% of the Gas limit for the above two transactions. For the last transaction of Transfer NFT, BNB's Gas level was only 1.9% with 57,003 (i.e., lowest) while the highest Gas level was recorded by Fantom and Polygon with 93.41% of Gas limit; while BNB and Celo use 79.17% and 69.8% of Gas limit.

#### D. Gas Price

Table IV shows the value of Gas for all four platforms. Specifically, BNB, Fantom, and Celo have the same Gas value in all three transactions with values of 10 Gwei (i.e. the highest of the three platforms), 3.5 Gwei, and 2.7 Gwei, respectively. Meanwhile, the Gas value of Polygon platform (i.e. MATIC) has the lowest value and fluctuates around 2.5 Gwei.

## VI. DISCUSSION

According to our observation, the transaction value depends on the market capitalization of the respective coin. The total market capitalization of the 4 platforms used in our review (i.e. BNB (Binance Smart Chain); MATIC (Polygon); FTM (Fantom); and CELO (Celo)) are \$50,959,673,206, respectively; \$7,652,386,190; \$486,510,485; and \$244,775,762.<sup>25</sup> This directly affects the coin value of that platform - although the number of coins issued at the time of system implementation also plays a

<sup>20</sup>Implementation of theme models our release at Nov-24-2022 07:42:20 AM +UTC

<sup>21</sup><https://testnet.bscscan.com/address/0x2ec70f233d91ade867259ff20c75f5c54e1ff008>

<sup>22</sup><https://mumbai.polygonscan.com/address/0xd9ee80d850ef3c4978dd0b099a45a559fd7c5ef4>

<sup>23</sup><https://testnet.ftmscan.com/address/0xd9ee80d850ef3c4978dd0b099a45a559fd7c5ef4>

<sup>24</sup><https://explorer.celo.org/alfajores/address/0xD9Ee80D850eF3C4978Dd0B099A45a559fD7c5EF4/transactions>

<sup>25</sup>Our observation time is 12:00PM - 11/26/2022.

TABLE IV. GAS PRICE

	Contracts Creation	Create NFT	Transfer NFT
BNB Smart Chain	0.00000001 BNB (10 Gwei)	0.00000001 BNB (10 Gwei)	0.00000001 BNB (10 Gwei)
Fantom	0.0000000035 FTM (3.5 Gwei)	0.0000000035 FTM (3.5 Gwei)	0.0000000035 FTM (3.5 Gwei)
Polygon	0.000000002500000012 MATIC (2.500000012 Gwei)	0.000000002500000016 MATIC (2.500000016 Gwei)	0.000000002500000016 MATIC (2.500000016 Gwei)
Celo	0.0000000026 CELO (Max Fee per Gas: 2.7 Gwei)	0.0000000026 CELO (Max Fee per Gas: 2.7 Gwei)	0.0000000026 CELO (Max Fee per Gas: 2.7 Gwei)

huge role. The total issuance of the four coins BNB, MATIC, FTM, and CELO is 163,276,974/163,276,974 coins, respectively; 8,868,740,690/10,000,000,000 coins; 2,541,152,731/3,175,000,000 coins and 473,376,178/1,000,000,000 coins. The value of the coin is conventionally based on the number of coins issued and the total market capitalization with a value of \$314.98; \$0.863099; \$0.1909; and \$0.528049 for BNB, MATIC, FTM, and CELO, respectively.

Based on the measurements and analysis in Section V, we have concluded that the proposed model deployed on Faltom brings many benefits related to system operating costs. In particular, generating and receiving NFTs has an almost zero fee (i.e. negligible). Also, the cost of creating contracts with transaction execution value is also very low (i.e. less than \$0.002).

In future work, we proceed to implement more complex methods/algorithms (ie, encryption and decryption) as well as more complex data structures to observe the costs for the respective transactions. Deploying the proposed model in a real environment is also a possible approach (i.e. implementing the recommendation system on the FTM mainnet). In our current analysis, we have not considered issues related to the privacy policy of users (i.e. access control [25], [20], dynamic policy [23], [24]) - a possible approach would be implemented in upcoming research activities. Finally, infrastructure-based approaches (i.e. gRPC [26], [27]; Microservices [28], [29]; Dynamic transmission messages [30] and Brokerless [31]) can be integrated into the model of us to increase user interaction (i.e. API-call-based approach).

## VII. CONCLUSION

Our research paper aims to expand the traditional delivery system (i.e. dependent on trusted third parties). For example, in the CoD model, the shipper receives the product deposit and payment from the buyer and then passes it back to the seller. The article highlights the risks in applying traditional models as well as the limitations of current approaches. Thereby, we propose a model that combines Blockchain technology, smart contracts, and NFT to eliminate the role of a trusted third party. Specifically, in our proposed model, the deposits of all three parties (i.e. seller, buyer, and carrier) are stored and noted as NFTs - the processing is stored in smart contracts. We have implemented the proposed model as proof-of-concept based on the Ethereum platform and Solidity language. We also deploy our smart contracts on four popular platforms supporting EVM (i.e. BNB, MATIC, FTM, CELO). Our analysis on all four platforms in all three transactions (i.e. contracts creation, NFT creation, NFT transfer) found that our proposed model is suitable for installation on the Fantom platform - having the lowest transaction costs compared to the other three platforms.

Possible development directions for our proposed model are presented in the discussion.

## ACKNOWLEDGMENT

This work was supported by Engineer Le Thanh Tuan and Dr. Ha Xuan Son during the process of brainstorming, implementation, and evaluation of the system. This work was also supported by the FPT University Cantho Campus, Vietnam.

## REFERENCES

- [1] X. S. Ha, H. T. Le, N. Metoui, and N. Duong-Trung, "Dem-cod: Novel access-control-based cash on delivery mechanism for decentralized marketplace," in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 2020, pp. 71–78.
- [2] K. L. Quoc, H. K. Vo, L. H. Huong, K. H. Gia, K. T. Dang, H. L. Van, N. H. Huu, T. N. Huyen, L. Van Cao Phu, D. N. T. Quoc *et al.*, "Sssb: An approach to insurance for cross-border exchange by using smart contracts," in *International Conference on Mobile Web and Intelligent Information Systems*. Springer, 2022, pp. 179–192.
- [3] M. Baygin, O. Yaman, N. Baygin, and M. Karakose, "A blockchain-based approach to smart cargo transportation using uhf rfid," *Expert Systems with Applications*, vol. 188, p. 116030, 2022.
- [4] A. G. Dragomir, T. Van Woensel, and K. F. Doerner, "The pickup and delivery problem with alternative locations and overlapping time windows," *Computers & Operations Research*, vol. 143, p. 105758, 2022.
- [5] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," *Decentralized Business Review*, p. 21260, 2008.
- [6] V. Buterin *et al.*, "A next-generation smart contract and decentralized application platform," *white paper*, vol. 3, no. 37, pp. 2–1, 2014.
- [7] Ethereum, "How our escrow smart contract works," 2022. [Online]. Available: <https://www.thenational.ae/business/technology/cash-on-delivery-the-biggest-obstacle-to-e-commerce-in-uae-and-region-1>
- [8] D. Sinha and S. R. Chowdhury, "Blockchain-based smart contract for international business—a framework," *Journal of Global Operations and Strategic Sourcing*, 2021.
- [9] Y. Madhwal, Y. Borbon-Galvez, N. Etemadi, Y. Yanovich, and A. Creazza, "Proof of delivery smart contract for performance measurements," *IEEE Access*, vol. 10, pp. 69 147–69 159, 2022.
- [10] J. Calbetó, A. Abareshi, N. Sriratanaviriyakul, M. Nkhoma, S. Pittayachawan, I. Ulhaq, F. Wandt, and H. X. Vo, "Lazada's last mile: Where no e-commerce company in vietnam had gone before," in *Informing and Information Technology Education Conference*, 2017.
- [11] "Two party contracts," 2022. [Online]. Available: <https://dappsforbeginners.wordpress.com/tutorials/two-party-contracts/>
- [12] OpenBazaar, "Truly decentralized, peer-to-peer ecommerce features," 2022. [Online]. Available: <https://openbazaar.org/features/>
- [13] P. V. R. P. Raj, S. K. Jauhar, M. Ramkumar, and S. Pratap, "Procurement, traceability and advance cash credit payment transactions in supply chain using blockchain smart contracts," *Computers & Industrial Engineering*, vol. 167, p. 108038, 2022.

- [14] Q. Meng, S. Hou, Z. Li, and S. Lu, "A uniform payment system for hyperledger fabric blockchain," in *Blockchain Technology and Application: Second CCF China Blockchain Conference, CBCC 2019, Chengdu, China, October 11–13, 2019, Revised Selected Papers 2*. Springer, 2020, pp. 121–130.
- [15] H. X. Son *et al.*, "Towards a mechanism for protecting seller's interest of cash on delivery by using smart contract in hyperledger," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 4, pp. 45–50, 2019.
- [16] H. T. Le, N. T. T. Le, N. N. Phien, and N. Duong-Trung, "Introducing multi shippers mechanism for decentralized cash on delivery system," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, 2019.
- [17] N. T. T. Le *et al.*, "Assuring non-fraudulent transactions in cash on delivery by introducing double smart contracts," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, pp. 677–684, 2019.
- [18] N. Duong-Trung *et al.*, "Multi-sessions mechanism for decentralized cash on delivery system," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 9, 2020.
- [19] X. S. Ha, T. H. Le, T. T. Phan, H. H. D. Nguyen, H. K. Vo, and N. Duong-Trung, "Scrutinizing trust and transparency in cash on delivery systems," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2020, pp. 214–227.
- [20] H. X. Son and N. M. Hoang, "A novel attribute-based access control system for fine-grained privacy protection," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 76–80.
- [21] N. M. Hoang and H. X. Son, "A dynamic solution for fine-grained policy conflict resolution," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 116–120.
- [22] Q. N. T. Thi, T. K. Dang, H. L. Van, and H. X. Son, "Using json to specify privacy preserving-enabled attribute-based access control policies," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2017, pp. 561–570.
- [23] S. H. Xuan *et al.*, "Rew-xac: an approach to rewriting request for elastic abac enforcement with dynamic policies," in *2016 International Conference on Advanced Computing and Applications (ACOMP)*. IEEE, 2016, pp. 25–31.
- [24] H. X. Son, T. K. Dang, and F. Massacci, "Rew-smt: a new approach for rewriting xacml request with dynamic big data security policies," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2017, pp. 501–515.
- [25] H. X. Son, M. H. Nguyen, H. K. Vo *et al.*, "Toward a privacy protection based on access control model in hybrid cloud for healthcare systems," in *International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on European Transnational Education (ICEUTE 2019)*. Springer, 2019, pp. 77–86.
- [26] N. T. T. Lam, H. X. Son, T. H. Le, T. A. Nguyen, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, and H. V. K. Nguyen, "Bmdd: A novel approach for iot platform (broker-less and microservice architecture, decentralized identity, and dynamic transmission messages)," *International Journal of Advanced Computer Science and Applications*, 2022.
- [27] L. N. T. Thanh *et al.*, "Toward a security iot platform with high rate transmission and low energy consumption," in *International Conference on Computational Science and its Applications*. Springer, 2021.
- [28] —, "Toward a unique iot network via single sign-on protocol and message queue," in *International Conference on Computer Information Systems and Industrial Management*. Springer, 2021.
- [29] L. N. T. Thanh, N. N. Phien, T. A. Nguyen, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, and H. X. Son, "Ioht-mba: An internet of healthcare things (ioht) platform based on microservice and brokerless architecture," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2021.0120768>
- [30] L. N. T. Thanh *et al.*, "Uip2sop: A unique iot network applying single sign-on and message queue protocol," *IJACSA*, vol. 12, no. 6, 2021.
- [31] —, "Sip-mba: A secure iot platform with brokerless and micro-service architecture," *International Journal of Advanced Computer Science and Applications*, 2021.

# Integrated Assessment of Teaching Efficacy: A Natural Language Processing Approach

Lalitha Manasa Chandrapati<sup>1</sup>, Dr. Ch.Koteswara Rao<sup>2</sup>  
School of Computer Science and Engineering, VIT-AP University  
Beside Secretariat, Amaravathi 522237  
Andhra Pradesh, India

**Abstract**—The most significant component in the education domain is evaluation. Apart from student evaluation, teacher evaluation plays a vital role in the colleges or universities. The implementation of a scientific and appropriate assessment method for enhancing teaching standards in educational institutions is absolutely essential. Conventional teacher assessment techniques have always been bounded to bias and injustice for single dimensional assessment criteria, biased scoring, and ineffective integration. In this regard, it is crucial to develop a specialized teacher evaluation assistant (TEA) system that integrates with some computational intelligence algorithms. This research concentrates on using Natural language processing (NLP) based techniques for empirically analysing teaching effectiveness. We develop a model in which a teacher is evaluated based on the content he delivers during a lecture. Two techniques are employed to evaluate teacher effectiveness using topic modelling and text clustering. By the application of topic modelling, an accuracy of 75% is achieved and text clustering achieved an accuracy of 80%. Thus, the method can effectively be deployed to assess and predict the effectiveness of a teacher's teaching.

**Keywords**—Teacher evaluation; topic modeling; clustering; Latent Dirichlet Allocation (LDA); K-means

## I. INTRODUCTION

The acute necessity for experts is an aspect of present socio-economic upheaval. It contributes significantly to the level of potential nurturing at the pedagogical stage. Universities and colleges should improve the monitoring of teacher efficacy in classroom and develop an appropriate evaluation system since the competence of teachers directly determines the teaching standards and the understanding of performance of pupils [1].

Teaching and teacher activity are intrinsically diverse and complicated entities. Teaching entails a multitude of activities and interactions both inside and beyond the classroom. Although the concept of evaluating quality of teacher seems simple, in reality it includes identifying, describing, gathering data on, and inferring from hundreds of complicated component variables. Teacher's competencies significantly influence quality of teaching and learner's knowledge reception; accordingly, universities or colleges should improve classroom supervision and develop a realistic scoring scheme [2]. Teachers can receive feedback on the implementation outcomes to reinforce or improve particular areas of their training and make sure that they finish their assignments within the allotted amount of time by monitoring and evaluating the quality of their student's learning [19]. To continually increase student's learning abilities and teacher's teaching abilities, a faultless,

complete, and appropriate teaching performance evaluation system must be established [3].

At this moment, the implementation of a quantitative assessment method is necessary. The conventional form of teaching assessment continues to be used in the majority of universities and colleges [4]. The conventional teaching assessment methodology has proven inadequate to match the instructional criteria in the effective instructional procedure [5]. However, the conventional assessment methods do have some drawbacks, such as the single dimension of the indexes and the absence of a fairly neutral evaluation criteria [6].

Deep-learning is used in various aspects such as object detection, deep auto-encoders [21], image-captioning [24], sentiment analysis [22,25], etc. With the continual advancement in the areas of Deep learning and NLP, it is relatively possible to monitor teaching efficacy of a teacher in real time [7,20]. Teacher evaluation assistance system plays a crucial part in enhancing the effectiveness of potential nurturing at the teaching level [2,23]. Besides that, the teacher will get an opportunity of being awarded by the organization relying on the TEA system's results. Several factors can be taken into consideration while evaluating a teacher. Sentiment [17], presentation abilities, speaking skills, and student reviews are some of the common features. In this paper, we propose a TEA system that presents the teacher effectiveness score which solely depends on the content that the teacher delivers irrespective of his body postures, voice, etc.

$$TUS \propto \text{bodypostures} + \text{gestures} + \text{voice} + \text{accent} + \text{content} \quad (1)$$

In the eq. (1), TUS represents teacher uniqueness score

The teacher uniqueness score (TUS) is a measure that describes how distinct a teacher is from the others. As shown in the eq. (1), TUS depends on the body postures, gestures, voice, language and content. We consider 'content' as the main feature.

TABLE I. FEATURE INTERPRETATION IN TEA

Features/Aspects	TEA	[8]	[9]
Text extraction	yes	no	no
Student reviews	no	yes	yes
Context based	yes	no	no
Clustering	yes	no	no
Topic Heterogeneity	yes	no	no

The Table I displays the main discrepancies between our findings and that of others. As previously stated, this approach



incorporates text extraction, whereas the other two do not. Student reviews are required for assessing the teacher for the analysis done in [8] and [9], however they are not considered in this study. This research is solely context based, which has not been presented in prior studies. In this work, the clustering technique is applied for teacher assessment, whereas [8] and [9] use deep neural networks, deep denoising autoencoders, and support vector regression models. Topic heterogeneity, that is, different topics are considered when acquiring the dataset, unlike in the prior works.

The contributions of this article includes the following:

- This research developed a novel teaching quality assessment technique for universities and colleges based on topic modelling and clustering techniques.
- The proposed system solely concentrates on the content delivered by the teacher while instructing a class unlike other research works.
- This system can be used to assess a teacher as well as determine 'how effective a speaker is' in any given video.

The rest of this paper is organised in the following way. Section II addresses significant contribution in the field. Section III provides the proposed methodology of this system, Section IV describes experimental details followed by the results and finally, the conclusion is given in Section V.

## II. PRIOR RESEARCH AND NOVELTY OF TEA

The key approach for cultivating talents for a new generation is the enhancement of teaching excellence in academia, where the most significant aspect being the improvement of the foundation of teaching quality administrative systems. An efficient teacher assessment system in academia can depict teaching accomplishments in real times. The research on teacher assessment techniques in advanced nations seems to be more established owing to the early emergence of the education sector. In their paper [8], a comprehensive network for assessing teacher performance in universities and colleges was developed. The assemble evaluation data sets for the network, were collected from three different groups: students, peers, and leaders through survey questionnaire. The network of Student-NET, Peer-NET and Leader-NET was individually trained upon every dataset to build the hypothetical correlations between various evaluation indexes and outcomes from three perspectives. An integrated network Integrated-NET was implemented to merge the evaluations from the preceding networks in order to familiarise multi-dimensional analyses. This research eventually developed and implemented an online teaching assessment system, relying on the SSM framework, with a user-friendly functionality.

Yu Liu [9] in their study, analysed the traits and existent concerns of teacher's teaching assessment along with the critical aspects and approaches of university teacher teaching assessment. They designed an efficient methodology, a distinctive blend of deep denoising autoencoder and support vector regression model to assess the teaching effectiveness of college and university teachers. The model was built on multiple hidden layers and executed multiple feature transformations throughout the unsupervised training stage to achieve the

reconstruction between the output data and the input data. The developed approach was assessed in order to highlight the method's exceptional efficacy in enhancing decision-making about teacher effectiveness and its ability to accurately evaluate and anticipate the excellence of university education.

A research framework for analysing Physical Education teaching was proposed by Yuansheng Zeng [10], in which the authors developed a model for assessing both the teaching ability of the teacher and student's learning impact relying on a hybrid technology of data mining and the hidden Markov model respectively. Eventually, the cumulative assessment score is obtained.

In academic contexts, teachers, students, graduates and employers are often surveyed and this survey questionnaire includes open ended questions whose analyzation requires large amount of time and workload. To overcome this challenge, Buenano-Fernandez et al. in their study [11] suggested the use of a comprehensive technique based on topic modelling and text network modelling, which enables researcher to extract significant data from surveys containing open-ended questions. By use of these assessments vital data is gathered with the intention of determining the level of satisfaction that the aforementioned entities have with institutional education procedures.

A unique approach for evaluating the impact of practical knowledge on educational accomplishments was indicated by M. M. Rahman et al., in their study [12]. A remarkable approach for retrieving latent features and corresponding association rules from a legitimate dataset is presented in this research. For data clustering, an unsupervised k-means clustering technique is utilised, followed by the frequent pattern-growth approach for association rule mining. Using this framework many significant and relevant features were derived that are strongly linked to the learner's activities. To analyse the association among pragmatic (e.g. programming, logical implementations, etc.) abilities and entire scholastic accomplishment, statistical aspects of students are assessed, and the associated findings are presented. Relying on the determined latent attributes, a range of significant recommendations are presented for students for each cluster. Furthermore, the empirical outcomes of this study can assist teachers in developing productive instructional strategies, assessing programmes with precise arrangements, and pinpointing student's academic inadequacies.

By merging DBSCAN and k-means algorithms [13], the authors developed an ensemble unsupervised clustering paradigm for assessing student's behavioural traits. The efficiency of the suggested methodology is assessed by performing research on six categories of behavioural data generated by students at a Beijing university and assess the associations among diverse behavioural traits and student's grade point averages (GPAs). Besides detecting aberrant behavioural trends, the conclusions drawn from experiment also detect conventional behavioural trends more reliably.

As shown in Table II the prior work had some limitations. As it is seen, the data collected primarily was based on only reviews by different groups in their research [8]. Apart from those expressions, content, etc. could have been considered to evaluate the teacher. In their research [9], the data was collected only from teaching procedure with students, would

TABLE II. SUMMARY OF RELATED WORK

Reference #	Methodology	Findings	Limitations
[8]	Implemented an online teaching assessment system, relying on the datasets gathered from three distinct groups of people. To achieve this, they used ANNs.	An accuracy of 98.59% was achieved on verification dataset. The results of the testing reveal that the system acts effectively and fits the criteria to encourage pedagogical improvement in universities and colleges.	The data obtained for evaluating a teacher is primarily based only on the reviews from students, peers and leaders. But they could have considered other sources like expressions, content, etc., for evaluating the teacher.
[9]	A novel framework was designed by combining deep denoising encoder and support vector regression model to evaluate quality of teaching.	In comparison to other models, this model attained an accuracy of 85.23%	It would have been better if they could have considered other parameters too for evaluation, rather than depending only on the data taken from teaching procedure with students.
[10]	Combination of data mining and hidden markov models to evaluate PE teaching effectiveness with regard to both teachers and students, in universities was proposed.	Attained better accuracy and also high computational efficiency compared to other models.	The training data should be large in size for the model to perform better and the computation time is high in this case.
[11]	LDA for topic modelling and text network modelling was used to glean relevant insight from questionnaires containing open-ended questions.	The deployment of this approach allows the optimization of effort and time required to adhere with the assessment of the text data produced by the open questions.	Equitably finite amount of data was considered in the proposed research and it is a challenging task to retrieve topics from concise text.
[12]	K-means clustering and FP-growth techniques were used to identify the correlation and affinity among practical skills and academic performance.	According to the study, it is concluded that stronger practical abilities have a good influence on academic success.	K value of the enhanced k-means clustering method may vary depending on the dataset and might result in yielding stronger or even worse outcomes for varied datasets.
[13]	Proposed the application of DBSCAN and K-means clustering to analyse the behavioural patterns of the students.	The outcomes of this research help in providing students, effective services and governance, like psychological assistance and educational counselling.	The issue of this research arises when applying the suggested method to multisource behavioural characteristics with high dimensions.

be better if other parameters are considered. In their study [10], large dataset should be taken for the model to predict better results. It is a challenging task to retrieve topics from concise text [11] considering finite amount of data. The limitation in the study [12] is that the k-value of the enhanced k-means method varies depending on the dataset resulting in yielding stronger or even worse outcomes for varied datasets. The issue of the research [13] arises when applying the suggested method to multisource behavioural characteristics with high dimensions.

Taking into account the shortcomings of previous studies, a novel approach for teacher evaluation is proposed. The study's novelty includes the teacher being evaluated solely on the content of his lectures, regardless of feedback or ratings.

### III. PROPOSED METHODOLOGY FOR TEA

For a college or university to develop and prosper, teaching standards must be of the highest calibre. So we try to implement a model that identify good teacher on their domain without human support. The detailed methodology is further discussed in this session.

#### A. A System level overview of TEA to evaluate teacher effectiveness

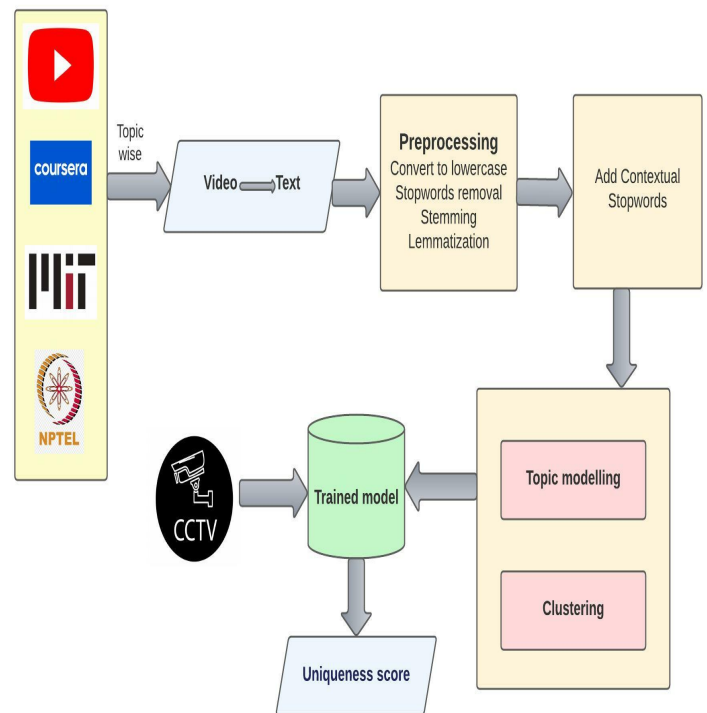


Fig. 1. Framework of the model to evaluate teacher

Fig. 1 depicts a schematic representation of the model for the TEA system. As illustrated in the Fig. 1, the steps involved in teacher evaluation are as follows:

- Videos spanning multiple sources such as YouTube, MIT courses, NPTEL, Coursera, and so on are considered.
- Using the IBM Watson and FFmpeg libraries, the acquired videos are transcribed to text and this text is treated as input.
- Pre-processing operations including converting to lowercase, stemming, lemmatization, stop word elimination, and so on are applied to the input text.
- Contextual stop words are added to the list.
- Topic modelling and Clustering approaches are employed on the pre-processed text and the trained model is stored.
- The lecture videos are acquired from the CCTVs installed in the classrooms, and are undergone through the trained model to provide a uniqueness score which describes “How effective a lecture of the teacher is?”

### B. Flowchart of the Proposed Model for Teacher Assessment

This research, which is based on topic modelling and clustering strategies, proposes a novel framework for teacher evaluation in universities and colleges, the flowchart of which is illustrated in Fig. 2.

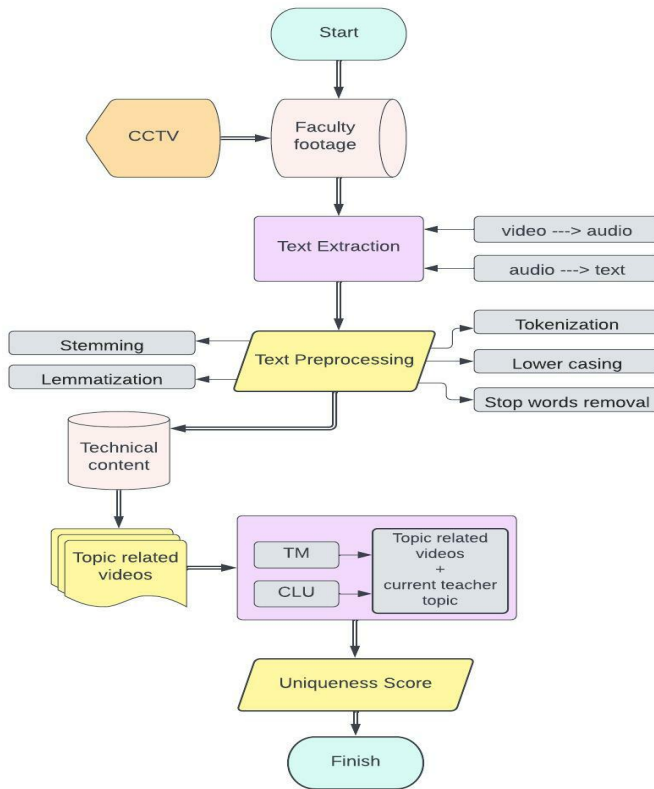


Fig. 2. Flow chart of the evaluation model for TEA based on topic modelling and clustering techniques

It can be seen from Fig. 2 that includes data required for the evaluation model is generated from Faculty footage

acquired of installed CCTVs at classrooms, then the generated data is undergone text extraction and text pre-processing steps. Finally, using both topic modelling and clustering techniques a uniqueness score is produced which evaluates the teacher effectiveness.

### C. Preparation of Data

As data acquired from CCTVs are videos, there is a need to transcript these videos to text format and also the transcribed text should be preprocessed.

1) *Text extraction*:: Converting videos to text can be accomplished in a variety of approaches, including utilising python, existing web APIs and soon. We transcript the videos using python FFmpeg and IBM Watson libraries here. The first phase includes conversion of video to audio using FFmpeg library. The second phase involves using IBM Watson library, which is SaaS provider for AI applications. The IBM Cloud offers a variety of solutions such as Text to Speech, Speech to Text, Natural Language Classifier, Language Translator, Visual Recognition, and so on. The Speech to Text service converts audio to text so that applications could use voice transcription features.

2) *Text preprocessing*: Text data is extensively available and is utilised to assess and solve business challenges. However, prior to actually using the data for research or prediction, it must be processed. Text preprocessing is used to prepare text data for model formulation. It is the first stage in any NLP project. Some of the preprocessing steps are: Removing punctuations like . , ! \$( ) \* % @, Removing Stop words, Removing URLs, Tokenization, Lower casing, Lemmatization, Stemming. Based on the dataset, we need to perform the appropriate preprocessing steps. For our dataset, we considered using Tokenization, Lowercasing, stop words removal, Lemmatization.

- **Tokenization**: The text is fragmented into small components in this stage. Based on our task specification, we can employ either sentence tokenization or word tokenization.
- **Lowercasing**: One of the most common preprocessing tasks is to convert the text to the same case, ideally lower case.
- **Stop words removal**: Stopwords are frequently used words that are eliminated from the text because they provide no relevance to the assessment. Those terms have little or zero significance. Apart from the existing list of stop words in NLTK library, we can modify the list by adding or eliminating terms depending on the scenario.
- **Stemming**: It involves stemming or reducing the words to it’s core form. For instance, programmer, programming, gets reduced to the word “program”. However, the drawback of stemming is that it separates the phrases, that the base form lacks meaning or is not reduced to a valid English word.
- **Lemmatization**: The difference between stemming and lemmatization is that it stems the term yet ensures that it retains its meaning.

#### D. Methodologies

For our research we have deployed two existing technologies to evaluate the teacher. The techniques used here are: Topic modelling and Clustering.

1) *Topic modelling-latent Dirichlet allocation*: Topic modelling is a type of unsupervised NLP that is used to depict a text document combining numerous topics that help describe the inherent content in a specific document. Analysis of text data is done using a model and clusters of words are generated on that dataset. Latent Dirichlet Allocation [14] is the most effectively used model for topic modeling approach [15].

In LDA, latent represents the underlying concepts in the data, while Dirichlet is a kind of distribution. The Dirichlet distribution is not same as normal distribution. In LDA [18], the topics for every document are allocated in the following manner:

- 1) For K number of predefined topics, in every document each word to a topic is arbitrarily configured.
- 2) For each document d:  
Determine the following for every word w in the document:
  - P(topic t / document d): The percentage of words in document d allocated to subject t.
  - P(word w / topic t): The percentage of assignments to topic t from words in w among all documents.
- 3) Given all previous words and their topic assignments, reassign topic T' to word w with probability  $p(t'/d)*p(w/t')$ .

The final phase is performed several times until we reach a stable condition in which the topic allocations do not vary any further. These topic allocations are then used to establish the percentage of subjects for each document.

2) *Clustering-K means*: The process of dividing a dataset into clusters is called Clustering. The idea is to divide the data so that elements within a single group are relatively same and those in different clusters are dissimilar. It defines how unlabelled data is categorized. In the context of using text data, it is Text Clustering, which involves different phases including Text pre-processing, feature extraction, Clustering. Text clustering uses machine learning and NLP to recognize and analyse textual data. The most well researched clustering algorithm is K Means [16], and it persistently tends to produce good results.

The main goal is to categorise given data collection into prespecified k number of distinct groups. The algorithm is divided into two phases: the initial step is to establish k centroids for each cluster, the following stage is to choose each point from the provided dataset and connect it to the closest centroid. To compute the distance among data points and centroids, Euclidean distance method is used extensively. Once all the elements are included in certain clusters, the first phase is accomplished. The revised centroids must be reassessed at this stage since the addition of new points could cause a variation in the centroids of the clusters. Once k new centroids are identified, a new association among the same data points and the closest new centroid is established, resulting in a

loop. As a consequence of this loop, the k centroid's locations may vary in a step-by-step process and this leads to a condition called clustering convergence which means the centroids stop drifting.

#### E. Our Proposed Methodology

1) *Evaluation parameters*: We consider normalized value of likes/views of a video as a parameter for evaluating a teacher. It is known that for a real time lecture footage, neither likes nor views exist. For this purpose, we deploy a model where the real time videos are compared to you tube videos taking NL/V into consideration. As a result, if the video needs to be uploaded to a web source, then the effectiveness score can be known prior to uploading thus making it easy to identify whether the video be success or not.

##### 2) Users of our system:

- Public domain(YouTube, MIT, Coursera, etc.)  
As mentioned earlier it would be easy to depict a video to be success or not using our proposed model. This can help the trainer to improve his presentation and provide better videos. Also, the effectiveness of the speaker in a video can also be depicted.
- College/University Management  
They can be benefited in varied ways like to assess a teacher without direct classroom intervention. A trainer can also be recognised and awarded by the management for his teaching abilities by application of this model.

## IV. VALIDATION OF TEA

### A. Dataset Collection

The major source of dataset for evaluating a teacher is the lecture footage acquired from classrooms. It is a challenging task to accumulate a significant amount of data based just on footage in a short span of time. As a result, we used YouTube videos to train the model, and once the model gives constructive results, it can be applied on real-time lecture footage.

Videos are converted to text format and the data is undergone through pre-processing techniques thus resulting in providing clean data. Topic modelling and clustering approaches, LDA and K-Means, respectively are deployed on the data to assess the teacher.

As shown in Table III, videos for each sample topic are acquired and processed according to our research. Videos are converted to text format and the data is undergone through pre-processing techniques thus resulting in providing clean data. Topic modelling and clustering approaches, LDA and K Means respectively are deployed on the data to assess the teacher. Likes and views of each video is taken and percentage of likes to views is calculated. As the videos are of varied durations, there is a need for normalizing the likes/views value. The normalization technique used here is minmax normalization. We contemplate NL/V as evaluation parameter, it is equal to 1 indicates that likes of the video are identical as the views of video. This depicts that the video is best as "who all viewed it, liked it". Hence, it can be deduced from the above said logic that  $NL/V = 1$  means the video is best of all.

TABLE III. DATASET

S.No.	Sample topic
1	What is an activation function?
2	Distinguish between AI vs ML
3	Explain Generations of computers
4	Explain GSM architecture
5	What is MQTT protocol?
6	What is a neural network?
7	What is virtualization?
8	What is cloud computing?
9	What is database?
10	What is data mining?
11	Differentiate between classification and regression
12	What is an embedded system?
13	Define gradient descent
14	Define internet of things
15	Introduction to machine learning
16	Explain OOPs concepts
17	What is an operating system?
18	Differentiate between supervised and unsupervised learning
19	What are the types of operating systems?
20	What are the types of neural networks?

B. Experimental Procedure

1) *Topic modelling*: Data pre-processing is performed first. After pre-processing the data, topic modelling is implemented on the data to extract significant topics. Word frequency count is calculated on the collection of words obtained for each topic. Frequent words with a threshold value of 4 or less are removed as more frequently occurred words in the corpus determines that the speaker tried to deliver the significant content by insisting those words. Average word count and average normalized count of the words are computed.

$$WC = \sum \text{frequency of words in topics} \quad (2)$$

$$AWC = \frac{\sum WC}{\text{Total no. of topics}} \quad (3)$$

2) *Clustering*: For clustering approach, the data is summarized using python code. K-Means clustering was deployed on the summarized data thus resulting in clusters. The optimal number of clusters is predicted by the usage of elbow method. Once the clusters were formed the participation ratio of every video across each cluster is calculated. Average clustering score for each video is also calculated. High average clustering score depicts the corresponding video being the best of all.

$$PR = \frac{\text{No. of sentences of a video in the cluster}}{\text{Total no. of sentences across the cluster}} \quad (4)$$

Where PR = Participation Ratio of video

$$ACS = \frac{\sum_{i=1}^k PR}{k} \quad (5)$$

Where ACS = Average clustering score, K = Total number of clusters

C. Results and Analysis

After applying topic modelling on each video of the dataset, topics are generated which are a collection of words. Word count and Average Word Count are calculated using the eq. (2) and (3), respectively across each topic for all the videos. Average word count is normalized using minmax normalization to obtain average normalized count values.

*Intertopic distance plot* The intertopic distance plot is also created using the topics that were generated. It presents a broad perspective of the topics and their difference from one another, whilst providing for a detailed analysis of the words most intrinsically correlated with each topic. The areas of the circles encode the relative prevalence of each topic. The horizontal bar chart is illustrated on the right whose bars indicate the distinct words which are most beneficial for comprehending the present topic of interest on the left.

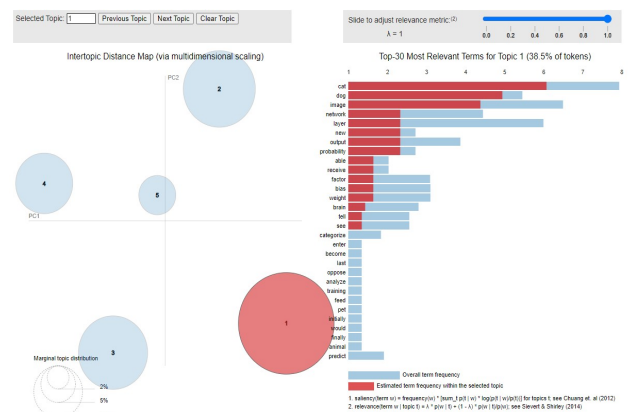


Fig. 3. Intertopic distance plot for neural network

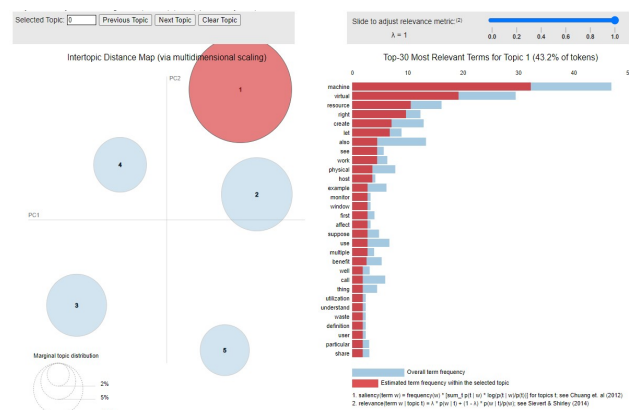


Fig. 4. Intertopic distance plot for virtualization

TABLE IV. STATISTICS OF TOPICS AFTER TOPIC MODELLING

TITLE	VIDEO-1	VIDEO-2	VIDEO-3	ANC(HIGH)
What is neural network?	NL/V=0 AC=31.4 ANC=0.44	NL/V=1 AC=41 ANC=0.47	NL/V=0.408 AC=38.6 ANC=0.45	YES
What is virtualization?	NL/V=0.57 AC=71 ANC=0.54	NL/V=1 AC=80.2 ANC=0.6	NL/V=0 AC=74.4 ANC=0.52	YES
What is database?	NL/V=1 AC=31.4 ANC=0.557	NL/V=0.329 AC=36.6 ANC=0.425	NL/V=0 AC=74.4 ANC=0.568	YES
What is cloud computing?	NL/V=0.67 AC=22 ANC=0.58	NL/V=1 AC=16.6 ANC=0.65	NL/V=0 AC=28.4 ANC=0.64	YES
What is data mining?	NL/V=0 AC=22.2 ANC=0.58	NL/V=0.414 AC=25.2 ANC=0.21	NL/V=1 AC=61.6 ANC=0.6	YES
Explain GSM architecture	NL/V=0 AC=24.4 ANC=0.408	NL/V=1 AC=36.8 ANC=0.48	NL/V=0.103 AC=62.6 ANC=0.38	YES
Distinguish between AI vs ML	NL/V=0 AC=22.2 ANC=0.26	NL/V=0.63 AC=36.4 ANC=0.6	NL/V=1 AC=38.6 ANC=0.51	NO
what is an activation function?	NL/V=0.85 AC=45.2 ANC=0.41	NL/V=1 AC=52.2 ANC=0.67	NL/V=0 AC=62 ANC=0.51	YES
What is MQTT protocol?	NL/V=1 AC=74.6 ANC=0.57	NL/V=0 AC=143 ANC=0.36	NL/V=0.06 AC=50.2 ANC=0.38	YES
Explain Generations of computers	NL/V=0.59 AC=42.4 ANC=0.44	NL/V=0 AC=28.6 ANC=0.487	NL/V=1 AC=35.8 ANC=0.54	YES
Differentiate between classification and regression	NL/V=0 AC=41.6 ANC=0.53	NL/V=1 AC=83.2 ANC=0.64	NL/V=0.74 AC=64.4 ANC=0.43	YES
What is an embedded system?	NL/V=0.75 AC=57.2 ANC=0.52	NL/V=0 AC=90.6 ANC=0.5	NL/V=1 AC=64.6 ANC=0.56	YES
Define gradient descent	NL/V=1 AC=95.6 ANC=0.45	NL/V=0.083 AC=130.6 ANC=0.55	NL/V=0 AC=108 ANC=0.65	NO
Define internet of things	NL/V=0 AC=28 ANC=0.46	NL/V=1 AC=34.8 ANC=0.48	NL/V=0.7 AC=27.8 ANC=0.38	YES
Introduction to machine learning	NL/V=0 AC=46.2 ANC=0.57	NL/V=0.33 AC=25.4 ANC=0.49	NL/V=1 AC=34.2 ANC=0.28	NO
Explain OOPs concepts	NL/V=0 AC=65 ANC=0.43	NL/V=0.89 AC=51.4 ANC=0.65	NL/V=1 AC=103.4 ANC=0.51	NO
What is an operating system?	NL/V=1 AC=41.6 ANC=0.43	NL/V=0.26 AC=48 ANC=0.53	NL/V=0 AC=29.6 ANC=0.475	NO
Differentiate between supervised and unsupervised learning	NL/V=0 AC=93.2 ANC=0.5	NL/V=1 AC=122.6 ANC=0.67	NL/V=0.46 AC=115.2 ANC=0.62	YES
What are the types of operating systems?	NL/V=0.83 AC=47.4 ANC=0.59	NL/V=0 AC=116.4 ANC=0.41	NL/V=1 AC=218.6 ANC=0.65	YES
What are the types of neural networks?	NL/V=0 AC=128.2 ANC=0.48	NL/V=1 AC=73.8 ANC=0.55	NL/V=0.088 AC=66 ANC=0.42	YES

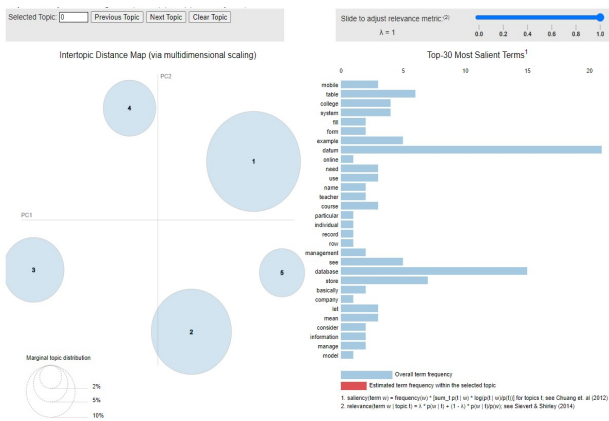


Fig. 5. Intertopic distance plot for database

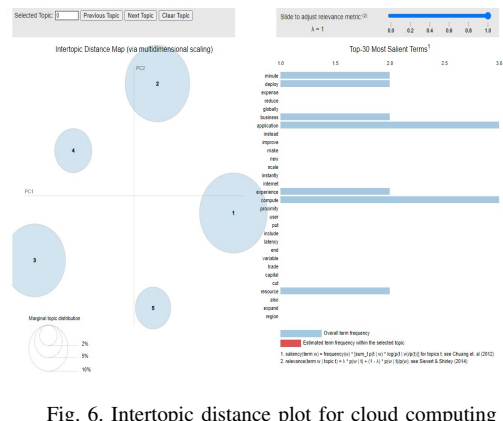


Fig. 6. Intertopic distance plot for cloud computing



Fig. 7. Intertopic distance plot for datamining

The images Fig. 3 through Fig. 7 depicts the intertopic distance plots of the five sample topics. It can be depicted that in distance plot the size of the circle represents each topic's population in the corpora. Each bubble depicts a distinct topic. The greater size of the bubble indicates the greater proportion of terms related to that topic. The greater the distance between the bubbles, the greater the degree to which they differ from one another. When there is more space between the bubbles, there is a huge differentiation among each of the individual bubbles. On the right side, the box plot represents the top thirty salient terms of each topic and their respective frequencies. The term frequency across the whole document is represented by blue bars and the red bars are categorized by the significance of the words within every topic.

As it is shown in Fig. 3 and Fig. 4, the bubble 1 is selected, thus red bars indicate the particular term's frequency inside the topic 1 for both respectively. For example, it can be seen in Fig. 4 that the term "machine" is the most frequently used for 48 times. The red bar depicts that the word is used for 32 times in that particular topic 1. Similarly in the plots Fig. 5, Fig. 6, Fig. 7, the frequency of each term in the corpus is shown by blue bars. The blue bars indicate that no topic is selected and hence, most widely spoken terms are presented. In Fig. 5, the most frequently used words are datum and database. As the plot is obtained from the sample topic that discusses about database, it is clear that the teacher is trying to induce the topic by concentrating on those words.

The Table IV represents the statistical measures calculated after acquiring topics by implementing LDA on the dataset. We calculated the term frequencies across the whole corpus for five topics over all the videos in the dataset and it is represented as WC, by using eq. (2). The average word count is obtained by usage of eq. (3) and can be identified as AC. ANC is the normalized value of AC. As seen in the Table IV ANC score is high for 15 sample topics out of 20 sample topics. Those fifteen instances whose ANC is high, it can be observed that their respective video's NL/V value is 1. Hence, a conclusion can be drawn from the observed pattern that ANC score is high for the videos whose NL/V value is equal to 1 which comprehends that the particular video is the best. The Fig. 8

represents a graph for the distribution of the sampled topics whose average normal count is high. The videos with high ANC are considered to be the best videos and it can be deduced that the teacher delivered the lecture in an effective manner compared to other two videos.

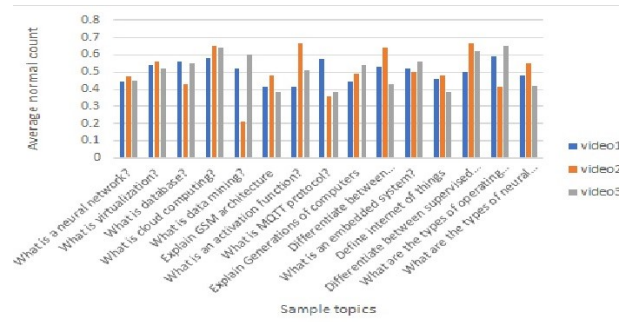


Fig. 8. Distribution plot of videos after topic modelling

TABLE V. TABLE OF CLUSTERING SCORE

Sample Videos	K	ACS(HIGH)	NL/V=1
what is an activation function	5	v2=0.401	v2
what is neural network	5	v6=0.128	v6
what is virtualization?	6	v3=0.18	v2
what is database?	8	v1=0.12	v1
what is cloud computing?	4	v4=0.2	v4
what is data mining?	8	v3=0.162	v3
Explain generations of computers	7	v9=0.121	v9
Explain OOPs concepts	7	v4=0.178	v10
Explain GSM architecture	6	v2=0.38	v2
what is MQTT protocol?	3	v2=0.37	v2

After applying clustering to the dataset, the clusters are formed and number of optimal clusters is identified by using elbow method. Using the eq. (4), participation ratio of the videos across the clusters for all the videos is calculated. ACS, average clustering score of all the sample topics is calculated using the eq. (5). The Table V depicts the number of optimum clusters for sample topics and their highest ACS score among the videos. It can be seen in Table V that 8 out of 10 sample topics have highest average clustering score of the corresponding video are same as the video having NL/V is equal to 1. Thus it can be concluded that the best videos of the sample topics after clustering have the highest average clustering score.

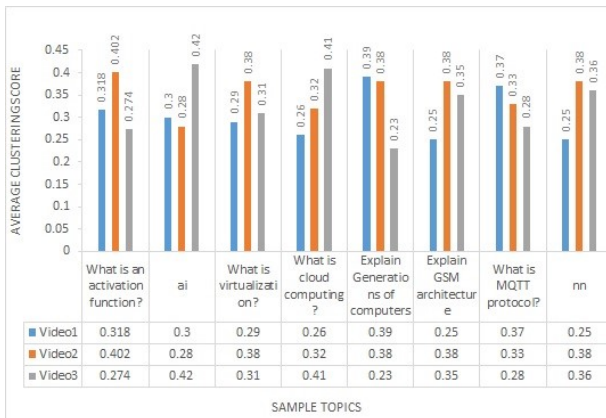


Fig. 9. Distribution plot of videos after clustering

The Fig. 9 shows the distribution plot of the sample topics after clustering technique been applied. It portrays the occurrences that have the highest average clustering score. Finally, We have drawn a conclusion from the patterns observed that by applying topic modelling and clustering methods, teacher evaluation can be performed based on the content and it got successful results with an accuracy of 75% and 80%, respectively.

## V. CONCLUSION

In this paper, LDA topic modelling and K-Means clustering methods are deployed on the dataset to evaluate a teacher's teaching effectiveness. The main feature for this research is considered as the content that the teacher delivers. Based on this context, videos from public domain are gathered for the dataset and applied the said models on the data resulting in obtaining an accuracy 75% and 80% of LDA, K-Means respectively. The utilization of two models is beneficial for acquiring effective results, like topic modelling consumes less computing time and clustering gives best accuracy. Finally, this paper provides a teacher evaluation system which is based on the content of the lecture distinct from traditional methods which are based on the reviews, emotion detection and so on.

## REFERENCES

- [1] Bain, Ken. What the best college teachers do. Harvard University Press, 2004.
- [2] Howe, Christine, et al. "Teacher-student dialogue during classroom teaching: Does it really impact on student outcomes?." *Journal of the Learning Sciences* 28.4-5 (2019): 462-512.
- [3] Chawinga, Winner Dominic. "Taking social media to a university classroom: teaching and learning using Twitter and blogs." *International Journal of Educational Technology in Higher Education* 14.1 (2017): 1-19.
- [4] Martinelli, Susan M., et al. "Results of a flipped classroom teaching approach in anesthesiology residents." *Journal of graduate medical education* 9.4 (2017): 485-490.
- [5] Kay, Robin, Thom MacDonald, and Maurice DiGiuseppe. "A comparison of lecture-based, active, and flipped classroom teaching approaches in higher education." *Journal of Computing in Higher Education* 31.3 (2019): 449-471.

- [6] Barrie, S., P. Ginns, and R. Symons. "Rewarding and recognising quality teaching and learning in higher education." (2007).
- [7] Joyce, Jeanette, Drew H. Gitomer, and Charles J. Iaconangelo. "Classroom assignments as measures of teaching quality." *Learning and Instruction* 54 (2018): 48-61.
- [8] Li, Guannan, et al. "Intelligent evaluation of teaching based on multi-networks integration." *International Journal of Cognitive Computing in Engineering* 1 (2020): 9-17.
- [9] Liu, Yu. "Evaluation Algorithm of Teaching Work Quality in Colleges and Universities Based on Deep Denoising Autoencoder Network." *Mobile Information Systems* 2021 (2021).
- [10] Zeng, Yuansheng. "Evaluation of physical education teaching quality in colleges based on the hybrid technology of data mining and hidden markov model." *International Journal of Emerging Technologies in Learning (iJET)* 15.1 (2020): 4-15.
- [11] Buenano-Fernandez, Diego, et al. "Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach." *IEEE Access* 8 (2020): 35318-35330.
- [12] Rahman, Md Mostafizer, et al. "Impact of practical skills on academic performance: A data-driven analysis." *IEEE Access* 9 (2021): 139975-139993.
- [13] Li, Xiaoyong, et al. "An unsupervised ensemble clustering approach for the analysis of student behavioral patterns." *Ieee Access* 9 (2021): 7076-7091.
- [14] Guo, Chonghui, Menglin Lu, and Wei Wei. "An improved LDA topic modeling method based on partition for medium and long texts." *Annals of Data Science* 8.2 (2021): 331-344.
- [15] Muchene, Leacky, and Wende Safari. "Two-stage topic modelling of scientific publications: A case study of University of Nairobi, Kenya." *PloS one* 16.1 (2021): e0243208.
- [16] Nazeer, KA Abdul, and M. P. Sebastian. "Improving the Accuracy and Efficiency of the k-means Clustering Algorithm." *Proceedings of the world congress on engineering*. Vol. 1. London, UK: Association of Engineers London, 2009.
- [17] Onan, Aytug, Serdar Korukoglu, and Hasan Bulut. "LDA-based Topic Modelling in Text Sentiment Classification: An Empirical Analysis." *Int. J. Comput. Linguistics Appl.* 7.1 (2016): 101-119.
- [18] Mutanga, Murimo Bethel, and Abdultaofeek Abayomi. "Tweeting on COVID-19 pandemic in South Africa: LDA-based topic modelling approach." *African Journal of Science, Technology, Innovation and Development* 14.1 (2022): 163-172.
- [19] Liang, Liqun, Qian Yin, and Chunrang Shi. "Exploring proper names online and its application in English teaching in university." *ASP Transactions on Computers* 1.1 (2021): 24-29.
- [20] Nava, Imelda, et al. "Measuring teaching quality of secondary mathematics and science residents: A classroom observation framework." *Journal of Teacher Education* 70.2 (2019): 139-154.
- [21] Gomed, Everton, Rodolfo Miranda de Barros, and Leonardo de Souza Mendes. "Deep auto encoders to adaptive E-learning recommender system." *Computers and Education: Artificial Intelligence* 2 (2021): 100009.
- [22] Pong-Inwong, Chakrit, and Konpusit Kaewmak. "Improved sentiment analysis for teaching evaluation using feature selection and voting ensemble learning integration." *2016 2nd IEEE international conference on computer and communications (ICCC)*. IEEE, 2016.
- [23] Ge, Dongjun, Xiaoyue Wang, and Jingting Liu. "A teaching quality evaluation model for preschool teachers based on deep learning." *International Journal of Emerging Technologies in Learning (iJET)* 16.3 (2021): 127-143.
- [24] Sirisha, Uddagiri, and Boleam Sai Chandana. "Semantic interdisciplinary evaluation of image captioning models." *Cogent Engineering* 9.1 (2022): 2104333.
- [25] Sirisha, Uddagiri, and Sai Chandana Boleam. "Aspect based Sentiment & Emotion Analysis with ROBERTa, LSTM." *International Journal of Advanced Computer Science and Applications* 13.11 (2022).



# An Effect Assessment System for Curriculum Ideology and Politics based on Students' Achievements in Chinese Engineering Education

Bo Wang<sup>1</sup>, Hailuo Yu<sup>2</sup>, Yusheng Sun<sup>3</sup>, Zhifeng Zhang<sup>4</sup>, Xiaoyun Qin<sup>5</sup>

Software Engineering College, Zhengzhou University of Light Industry  
Zhengzhou, China, 450000<sup>1,3,4</sup>

Zhengzhou University of Industry Technology  
Zhengzhou, China, 451150<sup>2</sup>

College of Material and Chemical Engineering  
Zhengzhou University of Light Industry, Zhengzhou, China, 450000<sup>5</sup>

**Abstract**—The curriculum ideological and political education (CIPE) has caused the attention of China's leaders and state departments in Chinese education, but its effect assessment is still an open issue should be address for the efficient and effective implementation of CIPE. The engineering education conception has been widely adopted in Chinese higher education, in recent years, due to its effectiveness. Therefore, in this paper, under the background of Chinese engineering education, we study the CIPE effect quantification. We propose a CIPE effect assessment system for higher education, and a CIPE effect quantitative method based on the achievements of graduation requirements for each student. The proposed system provides visualization information of achievements and CIPE effect for students and teachers. This helps students to locate themselves in their major learns, and teachers to continuously improve their teaching methods.

**Keywords**—Curriculum ideology and politics; assessment; engineering education; ideological and political education; outcomes-based education

## I. INTRODUCTION

Recent years, Chinese national leader and institutions, e.g., State Department and Education Ministry, put a high value on the curriculum ideological and political education (CIPE), to raise the ideological and ethical standards of Chinese students. They require all educators must pay equal attention to the professional ability training and the ideological and political education (IPE). In addition, IPE must be integrated into all aspects of education, instead of only relying on only a few IPE curricula. The Chinese Ministry of Education and other nine national-level government bodies jointly issued the official document of *Work Programme for Fully Pushing Forward the Construction of "Great Ideological and Political Course"* in July 2022 [1], which plans to fully mobilize social forces and resources for CIPE in all universities, middle and primary schools in China.

Therefore, many Chinese researcher study on how to organically integrating IPE into the professional education in various subjects, e.g., computer science [2], [3], [4], mathematics [5], [6], agronomy [7], and so on, to the isolated island phenomenon in IPE [8], and develop the world-class modern education with Chinese characteristics [9]. For example, Lu [10] applied new media network platform and block chain

technology to design a CIPE information exchange approach, to improve the sharing efficiency of IPE resources. This work also used the bad information screening technology to filter out negative information. Liu and Ni [11] exploited SWOT analysis method qualitatively analysing the inherent strengths and weaknesses as well as the external opportunities and challenges of the IPE politics employed by universities and colleges, currently, in HeiLongJiang province of China. Lin et al. [3] introduced a strategy for integrating IPE into the artificial intelligence course, and designed the IPE achievement goals of the course, based on the idea of outcome-based education (OBE). Zhang et al. [12] studied the combination of CIPE and Chinese engineering education accreditation, and integrated IPE into various educational phases including the educational goals, graduation requirements' indicators, the course system, and teaching designs. These works studied the integration strategy of CIPE in some aspects, but didn't consider to evaluate the effectiveness of their strategy in improving ideological and political achievements for students.

There are several works on the effect assessment of IPE education. Zhang [13] proposed to exploit industry education integration for CIPE education, which combined field practice and classroom teaching, to improve the mental health of students, by increasing the attraction of IPE education. Tang [14] studied the CIPE education reform for innovation and entrepreneurship education, to improve the self-efficacy of fresh graduates in their job applications. These works used grouping comparison to evaluate the effectiveness of CIPE strategies. Chen and Yu [15] used the principal component analysis technology to reduce the information redundancy of CIPE assessment criteria, and performed the weighted sum on the principal components to establish the evaluation model for CIPE education effect. Similarly, Zhang [16] employed the clustering algorithm to find the similarities of data indices, and built a CIPE evaluation data model by assigning the weight coefficient of each index class produced by factor analysis. Mo et al. [17] utilized triple helix model to combine the assessment data from three subjects, school, enterprise, and government, and established a fuzzy synthetic evaluating model for evaluating the CIPE effect. Ding et al. [18] used the back propagation neural network to evaluate the the CIPE teaching quality of teachers, with the data scored by students.

This work employed the traditional idea of teacher-centered instead of student- or learner-centered education. Zhang et al. [19] applied deep learn model for analysing the CIPE teaching quality, with multi-source data from on-line and classroom teaching, and produced the deficiency of teaching. Guo [20] designed an evaluation system of CIPE effect based on the authenticity evaluation theory for vocational education. All of these above works achieved the data through scoring subjective questions, leading to objectivity of CIPE effect assessment. Lv et al. [21] developed a data management system for IPE, where the evaluation of IPE effect is same to that of usual curricula, where teachers scoring.

To our best knowledge, there is lack of objective and quantitative evaluation method for CIPE effect. Existing works evaluated results of CIPE, ignoring the dynamic of the political and ideological status for each student. Therefore, in this paper, we study on the quantitative evaluation strategy for CIPE effect, based on the dynamic of achievements of students during their school years. The achievements of every student is the progress that the student reaches the graduation requirements specified in its specialty personnel training program (PTP). The graduation requirements are the main goal that undergraduates pursue in their four-year university or college life, in the context of engineering education with the idea of outcome-based education (OBE), in China. The graduation requirements are stable during students' school years. Thus, the fluctuates of students' achievements can reflect their IPE states. Such as, an improvement of achievements imply a better IPE effect (self-discipline) for a student. Our work complements to existing works, and helps to provide a more comprehensive assessment information of CIPE effect.

The rest of this paper is organized as follows. In Section II, we present preliminary knowledge on the Engineering Education. Section III illustrates our CIPE effect assessment system, and Section IV details the assessment method for quantifying the CIPE effect based on achievement data. In Section V, we illustrates some application cases for students and teachers, on the CIPE effect. In Section VI, we conclude this work and present our future work.

## II. PREPARATORY KNOWLEDGE ON ENGINEERING EDUCATION

In 2016, China became the 18th official member of Washington Accord, and the undergraduate engineering degree is mutual recognition among the countries joining the Washington Accord, for certified engineering specialties. This promotes the development of internationalization and modernization of Chinese higher education. Therefore, numerous Chinese universities actively apply for engineering education certification. In 2022, almost 2,000 university specialties have gained the certification from China Engineering Education Certification Association (CEEAA) [22].

### A. Engineering Education Conception

The concepts of engineering education are student-center, outcome orientation, and continuous improvement. There are one kind of eventual outcomes (cultivation goals) and two kinds of intermediate outcomes (graduation requirements and course objectives) in engineering education, As shown in

Fig. 1. The cultivation goals are the capacities that students will have after graduating about five years, which are designed and continuously improved based on the developments of social demands and expectations, the school and specialty orientations, etc. Graduation requirements are the capacities that students graduated with, by the college or university education. Graduation requirements are designed on the principle that they can support the reaching of cultivation goals. Course objectives are the capacities obtained by students through various courses directly. The accumulation of course objectives must support the achievements of all graduation requirements. And the course objectives are accomplished by corresponding teaching contents and methods.

As shown in Fig. 1, there are four continuous improvement circles or closed loops in engineering education, as illustrated in the followings.

- 1) The ultimate goal of engineering education is enabling students to achieve cultivation goals. Thus, the cultivation goals must be designed legitimately, as they decides the rationality of the whole PTP. Thus, when the cultivation goals are designed, their rationality must be assessed by related enterprise, industry, and education experts. When the assessed rationality is low, the cultivation goals are improved based on expert comments. These above steps are repeated for improving the cultivation goals until they are assessed to be rationality. The improving process is conducted periodically, as the social needs are changed with the world development. Usually, a major improvement is made in the cultivation goals every four years, and a minor one is made every two years, in Chinese higher schools.
- 2) Given the reasonable cultivation goals, the graduation requirements are designed to enabling students to have the ability to achieve these goals after graduation. Thus, the rationality of designed graduation requirements can be evaluated by the reaching degree of cultivation goals. If the reaching degree is low, the design of graduation requirements cannot support the achievements of cultivation goals, and thus need to be improved. These process is conducted when the cultivation goals are changed in the first continuous improvement circle or their evaluation results do not meet expectations.
- 3) Cultivation goals and graduation requirements are the high-level designs of PTP. Graduation requirements are the achievements that a student get in its about four years of engineering education, by various courses. Therefore, the course system must be built carefully, and designed course objectives need be equivalent to graduation requirements. After completing a course, a student accomplishes corresponding course objectives, and gets closer to corresponding graduation requirements. Thus, if the achievements of graduation requirements are evaluated to be poor for fresh graduates, there is a strong likelihood that course objectives are improper and they need to be improved to guarantee that students meet graduation requirements if they achieve all course objectives.
- 4) For each course objective, its accomplishment for students is implemented by reasonable designs and

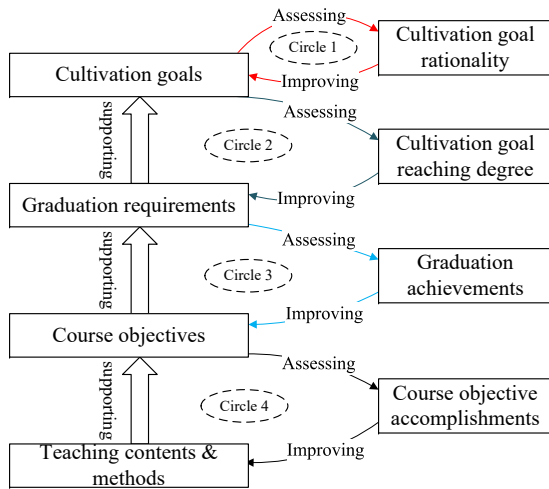


Fig. 1. The core concepts of engineering education

arrangements of course teaching contents and methods. When the accomplishment degree is low for a course objective, the corresponding teaching contents and methods need to be improved.

### B. Graduation Achievement Quantification

Given from CEEAA, certification standards show that there are 12 graduation requirements including engineering knowledge, problem analysis, solution design/develop, research, etc. To easily evaluating the achievement, each graduation requirement is decomposed into 2–5 smaller requirements, which can help students achieve the graduation requirement progressively and accumulatively.

Assuming that there are total  $N$  graduation requirements ( $g_i$ ) and  $M$  course objectives ( $c_j$ ). The relationship between graduation requirements and course objectives can be represented as a support matrix ( $\mathbf{W}$ ), where item  $w_{i,j}$  on line  $i$  column  $j$  is the support weight/degree of  $c_j$  to  $g_i$ . If  $c_j$  don't support  $g_i$ ,  $w_{i,j} = 0$ . For each graduation requirement, the achievement is quantified by the weighted sum of course objectives' accomplishment degrees, as shown in Eq. (1). Where  $r_i$  and  $o_j$  are respectively the achievement and accomplishment degrees of  $g_i$  and  $c_j$ , and they are both in the range from 0 to 1. And thus  $\sum_{j=1}^M w_{i,j} = 1$ , representing course objectives with non-zero weight jointly support the achieving for every graduation requirement. The accomplishment degree of each course objective is obtained by testing in the corresponding course.

$$r_i = \sum_{j=1}^M w_{i,j} \cdot o_j \quad (1)$$

All course objectives are tested only when near the graduation, as courses are run on all time periods (e.g. terms) of every student's school life. Thus, the graduation requirements are achieved incrementally for each student in its school life. In this paper, we focus on the fluctuation of graduation requirement achievements to quantify the CIPE effect for each student. Therefore, we normalizes graduation requirement

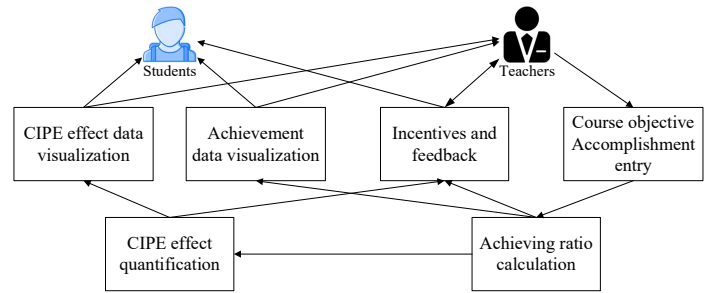


Fig. 2. The framework of CIPE effect assessment system

achievement degrees as achieving ratios in a same value range  $([0, 1])$  at any time, using following processes.

As course objectives are tested at different times, a student's school life can be divided into  $M$  time periods, by these test times. Without loss of generality, the test order is  $c_1, c_2, \dots, c_M$ . The  $j$ th time period,  $t_j$ , is the period between the times of testing  $c_{j-1}$  and  $c_j$ , and we can obtain the accomplishment degree of  $c_j$  at the end of  $t_j$ . Then, we define the achieving ratio ( $AR_{i,j}$ ) of  $g_i$  at  $t_j$  as the ratio of the current achievement degree of  $g_i$  to the full degree when  $o_{j'} = 1, \forall j' \leq j$ , as in Eq. (2).

$$AR_{i,j} = \frac{\sum_{j'=1}^j w_{i,j'} \cdot o_{j'}}{\sum_{j'=1}^j w_{i,j'}} \quad (2)$$

### III. CIPE EFFECT ASSESSMENT SYSTEM

The framework of CIPE effect assessment system is shown in Fig. 2. There are mainly six modules in the system, as detailed in the followings, respectively.

- M1. Course objective accomplishment entry module** provides entry interfaces (keyboarding, file import, etc.) for various courses' test data to teachers, and formats entered data into course objective accomplishment degree. For example, a course has two course objectives ( $c_1$  and  $c_2$ ), and its test data include scores of two assignments (80% and 75% for a student case), one middle exam (65%), and the final exam (77%). When the accomplishment of  $c_1$  is supported by these two assignments equally, and  $c_2$  by these two exams equally, the accomplishment degrees of  $c_1$  and  $c_2$  are respectively  $0.775 (\frac{80\%+75\%}{2})$  and  $0.71 (\frac{65\%+77\%}{2})$ , for the student case.
- M2. Achieving ratio calculation module** updates corresponding achieving ratios when the accomplishment degrees of some new courses objectives are calculated by M1 for some students, using Eq. (2).
- M3. CIPE effect quantification module** calculate the quantified data for the CIPE effect for each student, based on the fluctuation and the statistical information of achieving ratios. The quantification method is illustrated in Section IV.
- M4. Achievement data visualization module** uses some visualization tools to show the achievement data of graduation requirements for every student, including

the current achievement degrees and the fluctuations of achieving ratios. This can help the student to know its position in the specialty. For teachers, the module shows the statistical information of achievement data. In Section V, we will present some cases to illustrate the visualization of achievement data.

**M5. CIPE effect data visualization module** is similar to **M4**, which visualizes the CIPE effect data obtained from **M3**. Some cases will be illustrated in Section V.

**M6. Incentives and feedback module** recommends incentive strategies to students and teachers according to the historical achievement and CIPE effect data, and evaluates the effectiveness of an incentive strategy by testing whether there is a different before and after performing the incentive strategy. The recommendation of incentive strategies can be implemented by modern artificial intelligent algorithms, which is one of our future works.

#### IV. CIPE EFFECT QUANTITATIVE METHOD

In this paper, we design a CIPE effect quantitative metric based on the achieving ratio of each graduation requirement according to following principles.

- CIPE effect will be good if the achieving ratio is increased, and opposite when the achieving ratio is decreased.
- The small fluctuation of the achieving ratio can reflect the discipline of the student, and thus corresponds to a good CIPE effect.
- If the rank of a student is risen, the CIPE effect is likely to be good.

Thus, we construct the CIPE effect quantitative metric using following indicators.

- The increment of the two successive achieving ratios.  $\Delta_{i,j} = (AR_{i,j} - AR_{i,j-1})$  is the current increment at  $t_j$  for  $g_i$ .  $\Delta_{i,j}$  can be negative when the achieving ratio is decreased.
- The number of the achieving ratio increasing.  $I_{i,j} = \sum_{j'=1}^j (\Delta_{i,j'} > 0)$  represents the number for  $g_i$  at  $t_j$ , where  $(\Delta_{i,j'} > 0)$  returns 1 if  $\Delta_{i,j'} > 0$  and otherwise 0.
- The number of the achieving ratio decreasing,  $D_{i,j} = \sum_{j'=1}^j (\Delta_{i,j'} < 0)$ , similar to the previous indicator.
- The standard deviation of historical achieving ratios,  $\sigma_{i,j} = \sqrt{\sum_{j'=1}^j (AR_{i,j'} - \mu_{i,j})^2 / j}$ , where  $\mu_{i,j}$  is the average of achieving ratios, which is  $\sum_{j'=1}^j AR_{i,j'} / j$ . This indicator reflects the fluctuation degree of the achieving ratio.

The CIPE effect quantitative metric is defined by Eq. (3) based these above four indicators, for each graduation requirement, and the overall CIPE effect quantitative metric is defined

by Eq. (4). The greater values these metrics have, the better CIPE effect is.

$$e_{i,j} = \begin{cases} \frac{I_{i,j} \cdot (1 + \Delta_{i,j})}{\sigma_{i,j} \cdot D_{i,j}} & \text{if } \Delta_{i,j} > 0 \\ \frac{I_{i,j}}{\sigma_{i,j} \cdot D_{i,j} \cdot (1 - \Delta_{i,j})} & \text{else} \end{cases} \quad (3)$$

$$E_j = \sum_{i=1}^N \frac{e_{i,j}}{\max_{1 \leq i \leq N} e_{i,j}} \quad (4)$$

#### V. CASE STUDY

In this section, we first present cases for illustrating how students and teachers can be benefited from our CIPE effect assessment system by visualization information.

##### A. Visualization for a Student

For each student, our system first provides its latest achieving ratios for all graduation requirements by a combination chart, as shown in Fig. 3. As shown in the figure, the system provides not only its achieving ratios, but also the average and the best ones of all students, for the student. In addition, the chart highlights the achieving ratios not passing in red color. By comparing its achieving ratios with the average ones or the passing lines, the student can get a clear sight in its position and reasonably plan next goals.

In addition, the CIPE effect assessment system provides the dynamic changes of each graduation requirement achieving ratio for a student, in two forms, as shown in Fig. 4 and 5, respectively. Fig. 4 gives the achieving ratio for the first graduation requirement in all time periods. The system also provides the achieving ratios of others graduation requirements for the student, as same as Fig. 4. From Fig. 4, the student can see its achieving ratios, as well as the average and the best ones. In addition, Fig. 4 gives the suggest of the next goal for the student, according to its current achieving ratio. The suggested goal is set as the average achieving ratio of students with a higher grade than the current achieving ratio. For example, for a student, the current achieving ratio is 0.73. The average achieving ratio of students with achieving ratios between 0.8 and 0.9 is 0.84. Then the next goal is suggested to be 0.84. To give a more clear sight of the fluctuation of the achieving ratio, the system also provides the variations of the achieving ratio for each student, as shown in Fig. 5.

Based on the achieving ratio, the system calculates the overall CIPE effect by Eq. (4), and shows the result by a bar graph, as Fig. 6. By this graph, the system provides the overall CIPE effect, and the average and the best one for every student, which helps it to locate itself in its major learn and CIPE effect.

##### B. Visualization for a Teacher

For a teacher, the assessment system provides the statistical information on achieving ratio and CIPE effect by boxplot graphs, as Fig. 7 and 8, respectively. From these graphs, the teacher can see clearly the relative difference among achievements and CIPE effects of all graduation requirements. With the knowledge of which graduation requirement has poor or good achievements and CIPE effects statistically, the teacher can analysis the strengths and weaknesses of its teaching methods, and hasten the strengthening and make up for the weaknesses in the future.

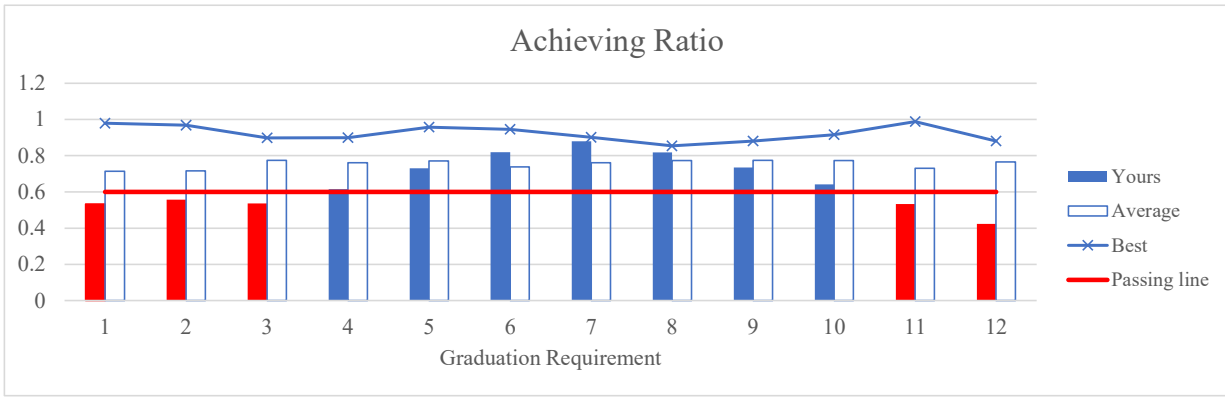


Fig. 3. The latest achieving ratios of a student

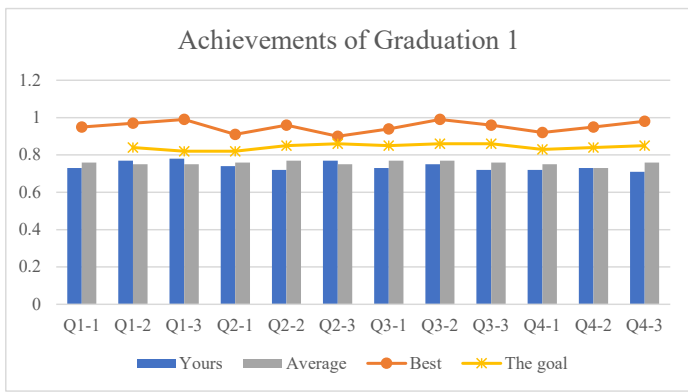


Fig. 4. The dynamical change of a graduation achievement for a student (Qa-b represents the time period, meaning the bth time period in the ath term)

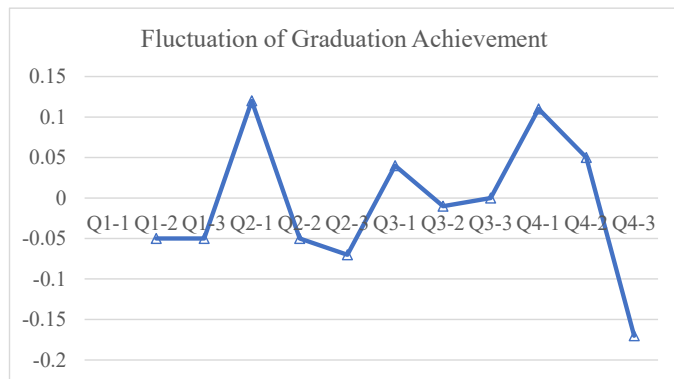


Fig. 5. The fluctuation of graduation achievement for a student

## VI. CONCLUSION

In this paper, we study on the CIPE effect assessment for Chinese higher education. We present a CIPE effect assessment system to manage the data of graduation achievements and the CIPE effect assessment of students, and propose a CIPE effect quantification method for the CIPE effect based on the graduation achievements. Our CIPE effect assessment system and method can be helpful for continuous improvements on achievements of students and teaching methods of teachers.

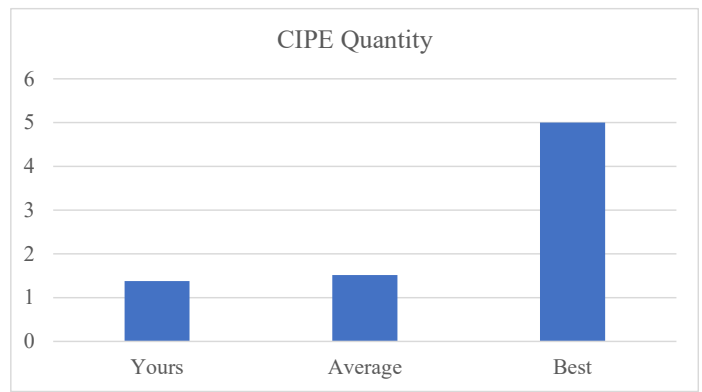


Fig. 6. The overall CIPE effect assessed value for a student

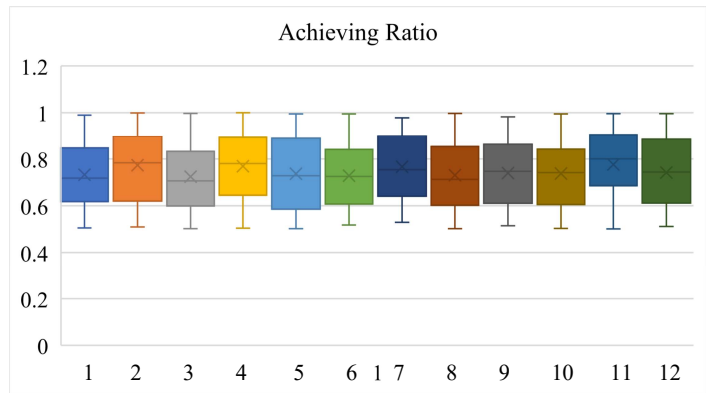


Fig. 7. The statistical information on achieving ratio for a teacher

To our best knowledge, this is the first attempt for quantify the CIPE effect based on objective data. In the future, we will study on the integration of our method with the subjective assessment for the CIPE effect, to provide more comprehensive knowledge for students and teachers.

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. The research was supported by Henan Higher Education Teaching Reform Research

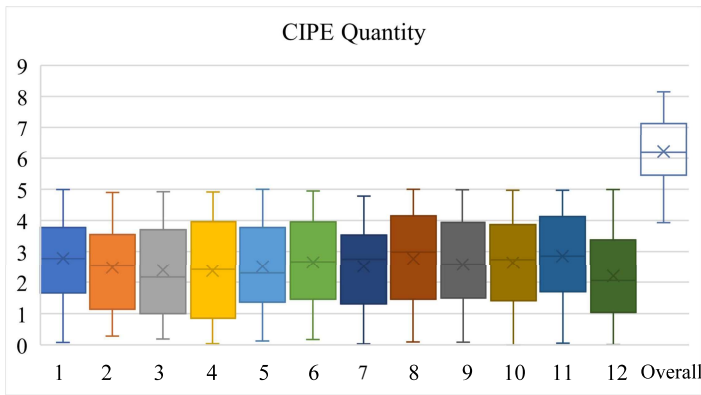


Fig. 8. The statistical information on CIPE assessment for a teacher

and Practice Project (Grant No. 2021SJGLX035, 2021SJGLX194, 2021SJGLX020).

#### REFERENCES

- [1] Chinese Education Ministry, et al., "Work Programme for Fully Pushing Forward the Construction of 'Great Ideological and Political Course'," [http://www.gov.cn/zhengce/zhengceku/2022-08/24/content\\_5706623.htm](http://www.gov.cn/zhengce/zhengceku/2022-08/24/content_5706623.htm), July 2022.
- [2] P. Zheng, X. Wang, and J. Li, "Exploration and practice of curriculum ideological and political construction reform - take "information security" course as an example," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 1–5, 2021.
- [3] X. Lin, Y. Wang, R. Zhang, T. Lin, J. Li, and X. Xue, "Ideological and political teaching reform: An introduction to artificial intelligence based on the obe concept," in *2022 11th International Conference on Educational and Information Technology (ICEIT)*, 2022, pp. 6–9.
- [4] Y. Li and H. Mao, "Study on machine learning applications in ideological and political education under the background of big data," *Scientific Programming*, vol. 2022, no. 3317876, pp. 1–9, 2022.
- [5] H. Zhu, Z. Zhang, and Y. Hua, "Teaching exploration and practice of advanced mathematics based on curriculum ideology and politics," *Curriculum and Teaching Methodology*, vol. 4, no. 4, pp. 74–78, 2021.
- [6] X. Huang and X. Wang, "A research on the strategy of integrating pbl with curriculum ideological and political elements - taking college mathematics as an example," *Journal of Contemporary Educational Research*, vol. 6, no. 7, pp. 1–5, 2022.
- [7] Y. Tian, T. Xu, J. Wang, and C. Chen, "Exploration on the reform path of ideological and political education for professional degree postgraduates - taking agricultural engineering and information technology field as an example," in *7th International Conference on Social Science and Higher Education (ICSSHE 2021)*. Atlantis Press, 2021, pp. 680–684.
- [8] Y. Guo, "The explanation of curriculum ideological and political theory under the construction of chinese road discourse system," *University Logistics Research*, vol. 242, no. 5, pp. 49–51, 2022.
- [9] N. Jin, "Modern education with Chinese characteristics and world level," *China Higher Education*, vol. 632, no. Z3, pp. 37–39, 2019, in Chinese.
- [10] H. Lu, "Design of ideological and political communication path of curriculum under the background of intelligent information of new media," *Mobile Information Systems*, vol. 2022, no. 4459877, pp. 1–11, 2022.
- [11] C. Liu and H. Ni, "Swot analysis based on heilongjiang university curriculum ideological and political research," in *Proceedings of the 4th International Seminar on Education Research and Social Science (ISERSS 2021)*. Atlantis Press, 2021, pp. 229–232.
- [12] Q. Zhang, Y. Zhao, and D. Xiang, "The fusion path of curriculum ideology and politics and engineering education accreditation," *Computer Education*, vol. 315, no. 3, pp. 76–80, 2021, in Chinese.
- [13] M. Zhang, "Research on the influence of curriculum ideological and political construction on college students' mental health under the background of industry education integration," *Psychiatria Danubina*, vol. 34, no. suppl 4, pp. 486–486, 2022.
- [14] J. Tang, "Research on the evaluation of students' employment psychology and self-efficacy in ideological and political curriculum reform under the background of innovation and entrepreneurship," *Psychiatria Danubina*, vol. 34, no. suppl 4, pp. 810–810, 2022.
- [15] X. kai Chen and J. Yu, "Evaluation model of physical education integrated ideology and politics based on principal component analysis," *Mobile Networks and Applications*, vol. 27, no. 3, p. 1240–1251, 2022.
- [16] Z. Zhang, "Modeling and analysis of quantitative evaluation of curriculum effectiveness based on clustering analysis algorithm," *Techniques of Automation and Applications*, vol. 41, no. 6, pp. 166–168+186, 2022, in Chinese.
- [17] Q. Mo, S. He, R. Tang, W. Liu, and W. Li, "Swot analysis based on heilongjiang university curriculum ideological and political research," in *2nd International Conference on Mechanical Engineering, Intelligent Manufacturing and Automation Technology (MEMAT 2022)*, 2022, pp. 1–7.
- [18] Y. Ding, W. Zeng, and Z. Ning, "Quality evaluation of ideological and political education in universities based on bp neural network," *Computational Intelligence and Neuroscience*, vol. 2022, no. 8336895, pp. 1–7, 2022.
- [19] B. Zhang, V. Velmayil, and V. Sivakumar, "A deep learning model for innovative evaluation of ideological and political learning," *Progress in Artificial Intelligence*, pp. 1–13, 2021, in Press.
- [20] G. Fang, "Construction of evaluation system of ideological and political education effect of vocational education curriculum based on authenticity evaluation theory," *Vocational and Technical Education*, vol. 43, no. 1, pp. 62–68, 2022, in Chinese.
- [21] Z. Lv, C. Zhou, and C. Zhou, "Design and implementation of ideological and political education evaluation system," *Journal of Physics: Conference Series*, vol. 1176, no. 4, pp. 1–7, 2019.
- [22] China Engineering Education Certification Association (CEECA), "Certification result enquiries," <https://www.ceea.org.cn/gcjyzyrzhx/gcjyzyrzhjlcx/index.html>, 2022.

# Navigation of Autonomous Vehicles using Reinforcement Learning with Generalized Advantage Estimation

Edwar Jacinto<sup>1</sup>, Fernando Martínez<sup>2</sup>, Fredy Martínez<sup>3</sup>  
Universidad Distrital, Francisco José de Caldas  
Bogotá D.C., Colombia

**Abstract**—This study proposes a reinforcement learning approach using Generalized Advantage Estimation (GAE) for autonomous vehicle navigation in complex environments. The method is based on the actor-critic framework, where the actor network predicts actions and the critic network estimates state values. GAE is used to compute the advantage of each action, which is then used to update the actor and critic networks. The approach was evaluated in a simulation of an autonomous vehicle navigating through challenging environments and it was found to effectively learn and improve navigation performance over time. The results suggest GAE as a promising direction for further research in autonomous vehicle navigation in complex environments.

**Keywords**—Actor-critic; autonomous vehicles; generalized advantage estimation; navigation; reinforcement learning

## I. INTRODUCTION

Autonomous vehicles have the potential to revolutionize transportation by reducing accidents, improving efficiency, and providing access to mobility for those who may not be able to drive [1], [2]. However, their deployment is limited by their ability to navigate complex environments, which can be affected by various factors such as weather, traffic, and pedestrians [3]. In addition, traditional approaches to autonomous navigation often rely on hand-designed rules or pre-defined maps, which may need to be revised to handle the variety and unpredictability of real-world environments [4], [5]. This can lead to poor performance and even accidents in complex scenarios.

From the point of view of robotic applications, one of the main problems in autonomous navigation is localizing the robot's position and orientation in the environment [6], [7]. This is typically done using sensors such as cameras, lasers, and inertial measurement units (IMUs), which provide noisy and partial observations of the environment [8], [9]. The robot must then fuse these observations with a map of the environment and estimate its pose using techniques such as Kalman filters, particle filters, or SLAM (simultaneous localization and mapping). However, these techniques can suffer from drift, ambiguity, and inconsistency, especially in dynamic or cluttered environments, leading to erroneous or uncertain pose estimates [10].

Another problem is planning and executing safe, efficient, and feasible trajectories [11], [12]. The robot must consider various constraints and objectives, such as avoiding collisions, respecting traffic rules, minimizing energy consumption, and

following a given path or mission [13]. This requires sophisticated algorithms that can reason about the robot's dynamics, kinematics, sensor models, and the environment's geometry, dynamics, and hazards [14]. These algorithms may include motion planners, path followers, and trajectory optimizers, which can be implemented using sampling-based planning, optimal control, and reinforcement learning techniques [15]. However, these techniques can be computationally intensive and may only sometimes find a solution, especially in complex or changing environments.

A third problem is adapting and learning from the environment [16]. The robot must learn and generalize from past experiences and observations to improve its performance, robustness, and flexibility [17]. This requires deep learning, transfer learning, and meta-learning, enabling the robot to learn features, models, and policies from data and transfer them to new tasks or situations [18], [19]. However, these techniques require large amounts of data and computation and may suffer from overfitting, generalization error, and sample efficiency.

To address this problem, a reinforcement learning approach using Generalized Advantage Estimation (GAE) to enable autonomous vehicles to learn to navigate complex environments is proposed [20], [21]. Reinforcement learning enables agents to learn by interacting with their environment and receiving feedback as rewards or penalties [22]. It has been successfully applied to various problems, including autonomous navigation [23]. However, traditional reinforcement learning approaches often require many interactions with the environment to learn effectively, which can be impractical in real-time scenarios such as autonomous navigation.

## II. BACKGROUND

GAE is a method for estimating the advantage of each action in a reinforcement learning algorithm, which is used to update the policy [24]. It was developed to address the problem of high variance in traditional reinforcement learning approaches, which can lead to slow learning and poor performance. GAE uses a linear combination of the value function and the reward to compute the advantage, which helps to reduce the variance and accelerate learning. It is effective in various environments, including robotic manipulation tasks [25].

To this end [26] propose Observational Imitation Learning (OIL), a novel imitation learning variant that supports online

training and automatic selection of optimal behavior by observing multiple imperfect teachers. [27] describe a generic navigation algorithm that uses data from sensors onboard the drone to guide the drone to the site of the problem. [28] propose a two-stage reinforcement learning (RL) based multi-UAV collision avoidance approach without explicitly modeling the uncertainty and noise in the environment. It is particularly an arduous task when handling multi-agent systems where the delay of one agent could spread to other agents. To resolve this problem [29] propose a novel framework to deal with delays as well as the non-stationary training issue of multi-agent tasks with model-free deep reinforcement learning. MIDAS uses an attention mechanism to handle an arbitrary number of other agents and includes a “driver-type” parameter to learn a single policy that aims of [30]. A neural network-based reactive controller is proposed for a quadrotor to fly autonomously in an unknown outdoor environment [31]. [32] aim to combine cloud robotics technologies with deep reinforcement learning to build a distributed training architecture and accelerate the learning procedure of autonomous systems. A deep reinforcement learning-based UANO (USVs autonomous navigation and obstacle avoidance) method is proposed [33]. Nowadays, modern Deep-RL can be successfully applied to solve a wide range of complex decision-making tasks for many types of vehicles. Based on this context [34] propose the use of Deep-RL to perform autonomous mapless navigation for Hybrid Unmanned Aerial Underwater Vehicles (HUAUVs), robots that can operate in both, air or water media. Other influential work includes [35].

This work proposes a reinforcement learning approach using GAE for autonomous vehicles navigating complex environments. Our method is based on the actor-critic framework, where the actor-network predicts the actions to take, and the critic network estimates the value of each state. GAE is used to compute the advantage of each action, which is applied to update the actor and critic networks. Our method is evaluated on a simulation of an autonomous vehicle navigating through a series of challenging environments and show that it can learn to navigate effectively and improve its performance over time. Our results demonstrate the potential of GAE for enabling autonomous vehicles to navigate through complex environments and suggest that it could be a promising direction for further research.

### III. PROBLEM STATEMENT

Autonomous vehicles have the potential to revolutionize transportation, but their ability to navigate complex environments is crucial for their practical deployment. Unfortunately, traditional approaches to autonomous navigation often rely on hand-designed rules or pre-defined maps, which may need to be revised to handle the variety and unpredictability of real-world environments. This can lead to poor performance and even accidents in complex scenarios.

Reinforcement learning is a promising approach for enabling autonomous vehicles to learn to navigate through complex environments. However, traditional reinforcement learning approaches often require many interactions with the environment to learn effectively, which can be impractical in real-time scenarios such as autonomous navigation. To address this problem, the use Generalized Advantage Estimation (GAE) to

reduce the variance and accelerate learning in reinforcement learning for autonomous vehicles is proposed.

This research aims to develop a reinforcement learning approach using GAE for autonomous vehicles navigating through complex environments and evaluate its performance on a simulation of an autonomous vehicle navigating through a series of challenging environments. The aim is to demonstrate that our method can learn to navigate effectively and improve its performance over time and to show that GAE has the potential to be a promising direction for further research in this area.

A simplified example to illustrate the basic ideas of using GAE to enable an autonomous vehicle to navigate through a complex environment can be seen in Algorithm 1. In practice, additional factors such as perception, planning, and dealing with real-world constraints and uncertainties must be considered. In this example, the environment is represented by the *Environment* class, which simulates the vehicle’s interactions with the environment. The *actor\_network* and *critic\_network* are neural networks that predict the actions to take and the value of each state, respectively. The Adam optimizer is used to update the networks based on the loss.

The algorithm’s main loop runs through the episodes, where an episode represents one complete run through the environment. Within each episode, the algorithm runs through the timesteps, where a timestep represents one action taken by the vehicle. At each time step, the algorithm predicts the *action* and the *value* using the actor and critic networks, takes action, and observes the next *state*, *reward*, and *done* flag. The *reward* and the *value* are stored in buffers, and the state is updated.

When the episode is complete, the algorithm computes the advantage using the GAE algorithm. It does this by first predicting the *value* of the final state using the critic network and then using this value along with the rewards and values from the episode to compute the returns using the *compute\_gae* function. The difference between *returns* and *values* then defines the *advantages*.

Then actor and critic losses are calculated. Actor loss is a measure of the quality of the action selection. It is computed using the log probability of the action taken, as predicted by the actor-network, and the advantage, as estimated by the GAE. The advantage represents the excess reward obtained from an action over the baseline value, and it reflects the relative importance of the action in the long run. The log probability represents the confidence of the actor-network in the action and reflects the risk of the action in the short run. The actor loss is defined as the negative dot product of these two quantities, which indicates the trade-off between exploration and exploitation. The actor loss is minimized during training to improve the action selection of the robot. Mathematically, the actor loss is defined as (Eq. 1):

$$actor\_loss = -mean(log\_probs \times advantages) \quad (1)$$

where *log\_probs* is a tensor of log probabilities, and *advantages* is a tensor of advantages.



---

**Algorithm 1** Pseudocode for Reinforcement Learning Algorithm using GAE

---

```
1: # Initialize the actor and critic networks
2: actor = Actor_network()
3: critic = critic_network()
4:
5: # Initialize the optimizer and the GAE parameters
6: optimizer = Adam(actor.parameters(), lr=learning_rate)
7: gamma = 0.99 # Discount factor for future rewards
8: beta = 0.95 # Weight factor for the advantage
9:
10: # Loop through the episodes
11: for episode in range(num_episodes): do
12:     # Initialize the environment and the state
13:     env = Environment()
14:     state = env.reset()
15:
16:     # Initialize the reward, value, and advantage buffers
17:     rewards = []
18:     values = []
19:     advantages = []
20:
21:     # Loop through the timesteps
22:     while not env.done: do
23:         # Predict the action and the value using
24:         # the actor and critic networks
25:         action = actor(state)
26:         value = critic(state)
27:
28:         # Take the action and observe the next state,
29:         # reward, and done flag
30:         next_state, reward, done = env.step(action)
31:
32:         # Store the reward and the value
33:         rewards.append(reward)
34:         values.append(value)
35:
36:         # Update the state
37:         state = next_state
38:     end while
39:     # Compute the advantage using GAE
40:     next_value = critic(state)
41:     returns = compute_gae(next_value, rewards)
42:     returns = compute_gae(values, gamma, beta)
43:     advantages = returns - values
44:
45:     # Compute the actor and critic losses
46:     actor_loss = (-log_probs * advantages).mean()
47:     critic_loss = advantages.pow(2).mean()
48:
49:     # Backpropagate the losses and update the actor
50:     # and critic networks
51:     loss = actor_loss + critic_loss
52:     optimizer.zero_grad()
53:     loss.backward()
54:     optimizer.step()
55: end for
```

---

The critic loss is a measure of the quality of the value estimation. It is computed using the advantage, as estimated by the GAE, and the value, as predicted by the critic network. The advantage represents the excess reward obtained from a sequence of actions over the baseline value, and it reflects the relative importance of the actions in the long run. The value represents the expected reward obtained from a state or action, and it reflects the long-term potential of the state or action. The critic loss is defined as the mean squared error between these two quantities, which indicates the deviation of the value from the true advantage. The critic loss is minimized during training to improve the value estimation of the robot. Mathematically, the critic loss is defined as (Eq. 2):

$$critic\_loss = mean(advantages.pow(2)) \quad (2)$$

where advantages is a tensor of advantages.

Finally, losses are backpropagated, and the model weights are updated. The losses are backpropagated through the model's computation graph to compute the model weights' gradients concerning the losses. The gradients are accumulated over an episode's timesteps and are used to update the model weights using the optimizer algorithm. The optimizer algorithm performs stochastic gradient descent on the model weights, using the gradients as the updated direction and the learning rate as the updated step size.

#### IV. METHODS

The algorithm is developed for a small autonomous robot under the following considerations that replicate the functional characteristics of our working platform (the robot could move to any of the four adjacent cells in the grid by selecting one of the following actions: up, down, left, or right). [36]:

- The robot has a state space of size four, representing the state of the robot in the environment.
- The robot has an action space of size two, which represents the actions that the robot can take in the environment.
- The robot can be rewarded for performing a specific action in the environment.
- The robot's state can change after taking action in the environment.
- The robot's episode can be terminated (done flag set to True) based on a particular environmental condition.
- The robot has an actor-network, which takes in the current state of the robot and outputs a probability distribution over the possible actions.
- The robot has a critic-network, which takes in the current state of the robot and outputs a value estimate for the current state.

Python and PyTorch are used for the implementation. The first step consists of importing the libraries and defining algorithm parameters, such as the number of episodes (1000) and the learning rate (0.001). Next, the actor and critic networks is

defined using PyTorch. The actor-network is a simple feedforward neural network with four input units (corresponding to the state size), two output units (corresponding to the action size), and two hidden layers with eight units each. The critic network is also a simple feedforward neural network with four input units (corresponding to the state size) and one output unit (corresponding to the value estimate). Both networks use ReLU activation functions.

Then, a function called `compute_gae()` is defined to compute the generalized advantage estimate for a given set of rewards, values, and discount factor. This function takes in the following value estimate, the rewards, the values, the discount factor, and the weight factor for the advantage, and returns the generalized advantage estimates.

The next step would be to define the environment with which the robot will interact. This can be done by creating a class for the environment, which should have methods for resetting the environment, stepping through the environment with an action, and checking if the episode is done. The environment should also have necessary attributes, such as the state and reward at each timestep and other relevant information. For example, in the case of a small robot navigating through a maze, the environment may have a 2D grid representation of the maze, with the state being the current position of the robot and the reward being based on the distance to the goal.

Finally, the algorithm's main loop is implemented, consisting of running through a series of episodes. In each episode, the environment is restarted, and the state is initialized. At each time step, the actor and critic networks are used to predict the action and the value, respectively, and the action is performed on the environment. The reward and value are stored in buffers. Once the episode is terminated (either by a termination condition or by reaching a maximum number of timesteps), the GAE algorithm is used to compute the rewards. For this purpose, the function `compute_gae()` is used, which takes the following value (predicted by the critical network), the rewards, the values (predicted by the critical network at each timestep), the discount factor (*gamma*) and the weight factor (*beta*). In the end, it returns the calculated payoffs. The losses of the actor and the critic are calculated using the advantages and the stored probabilities and values. These losses are backpropagated through the networks, and the optimizer updates the networks. This process is repeated for the specified number of episodes, updating the actor and critic networks at each iteration based on the observed rewards and advantages.

## V. RESULTS

The experimental setup was designed entirely in Python 3.8.16 (GCC 7.5.0) and respected the motion constraints of our ARMOS TurtleBot robot as well as the experimental navigation environment. The robot was modeled as an agent of the environment. The Deep-RL algorithm with GEA was also implemented with Python and the PyTorch 1.13.0 library.

For all training episodes, the initial position of the vehicle was set at the origin of the environment (lower left corner), with coordinates (0.0, 0.0), with its front pointing towards the positive x-axis. In each episode, a randomly generated target

point (final navigation destination) was created in the environment that the agent had to reach. In case of encountering an obstacle, or a boundary of the environment, the agent turns its position at a random angle and continues to move forward. The episode ends if the vehicle has reached a total of 1000 steps in that episode. Reaching the target point does not terminate the episode.

The Adam optimizer was used to train the actor and critical neural networks, with a learning rate of 0.001 for each method. In both instances, a 256-piece minibatch size was chosen. According to the reward values the agents attained, a training cap of 1000 episodes was established. In order to ensure a robust and effective learning algorithm, an exploration rate of 0.5 was employed for the actor network and 0.2 for the critic network.

The reward value of one of the tests developed can be seen in Fig. 1. It can be observed that it takes on average a little more than 200 episodes to start learning the task, but after this stage, it increases almost steadily the reward value until it reaches close to the maximum saturation value.

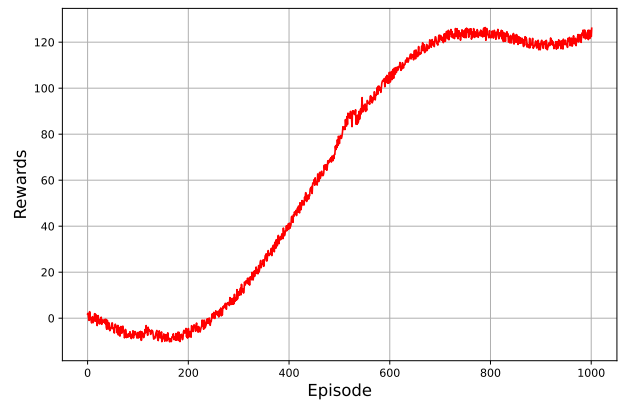


Fig. 1. Moving of the reward over 1000 episodes of the training

Our algorithm has been extensively validated and has consistently demonstrated accurate and effective agent behavior in the environment. The agents were able to learn the features of the environment and generalize their knowledge to successfully navigate without the use of external maps. In addition, they were able to adapt and learn how to avoid obstacles to reach their target destination. These results demonstrate the effectiveness of our algorithm in training intelligent agents to navigate complex environments in a lifelike simulation. This has important practical implications as it shows that our approach can be used to train agents for real-world navigation tasks without the need for external maps. Overall, our findings highlight the potential of our algorithm as a powerful tool for comprehending and learning complex behaviors in simulated environments.

## VI. DISCUSSION

The results of this study demonstrate the effectiveness of using a Deep-RL algorithm with GEA for navigation tasks in a simulated environment for a TurtleBot robot. The experimental setup, implemented in Python and utilizing the PyTorch library, successfully modeled the robot as an agent of the environment

and respected the motion constraints of the TurtleBot. The Adam optimizer was used to train the actor and critic neural networks, with a learning rate of 0.001 and a minibatch size of 256. The exploration rate was set to 0.5 for the actor network and 0.2 for the critic network.

The results show that the agent was able to effectively navigate to the randomly generated target points in the environment, while also avoiding obstacles and boundaries. It was observed that it took on average a little more than 200 episodes for the agent to start learning the task, but after this stage, the reward value increases almost steadily until it reaches close to the maximum saturation value. This suggests that the agent is able to learn the navigation task in a relatively short amount of time and perform well in the environment.

One limitation of this study is that the experiments were conducted in a simulated environment, which may not fully reflect the complexities of a real-world environment. Additionally, the agent's decision-making process was based on a predetermined set of rules, which may not be optimal in all situations. However, the study's results are still a promising step towards the development of intelligent robots that can navigate autonomously in complex environments.

In conclusion, this study has successfully demonstrated the effectiveness of using a Deep-RL algorithm with GEA for navigation tasks in a simulated environment for a TurtleBot robot. The results show that the agent is able to learn the navigation task in a relatively short amount of time and perform well in the environment. Future work can further improve the algorithm's performance by incorporating more realistic environments, and exploring other decision-making methods.

## VII. CONCLUSION

In this paper, a reinforcement learning algorithm for a small robot using generalized advantage estimation (GAE) is presented. The algorithm was implemented in Python and was designed to enable the robot to navigate and interact with its environment to maximize its reward. The robot was designed to operate in a simple 2D environment. The environment consisted of a grid of cells, each of which could be occupied by the robot or contain an obstacle. The robot could move to any of the four adjacent cells in the grid by selecting one of the following actions: up, down, left, or right. At each timestep, the robot received a reward based on its current position and the presence or absence of obstacles in its environment. To enable the robot to learn how to navigate its environment and maximize its reward, a reinforcement learning algorithm using GAE is implemented. The algorithm consisted of two key components: an actor-network and a critic network. The actor network was responsible for predicting the action the robot should take at each timestep based on the current state of the environment. The critic network was responsible for predicting the value of each state, which was used to estimate the expected future reward of each action. A combination of supervised learning and reinforcement learning to train the actor and critic networks is used. Specifically, the actor-network to predict the action the robot should take at each timestep based on the current state of the environment is used. The actor-network was trained using supervised learning,

with the input being the current state of the environment and the output being the predicted action. The critic network was trained using reinforcement learning, with the input being the current state of the environment and the output being the predicted value.

The algorithm was run for a series of episodes, resetting the environment and initializing the state at the beginning of each episode. At each timestep, the actor and critic networks predict the action and value, respectively, and the action is performed on the environment. The reward and value are stored in buffers. Once the episode is terminated, the GAE algorithm calculates the payoffs, and the losses for the actor and critic networks are calculated. These losses are backpropagated and used to update the networks. The results of our experiments show that the robot can successfully navigate through the environment and reach the goal state, with the average reward increasing throughout training. It was observed that the losses of actor and critic networks decrease as training progresses, indicating that the networks are learning effectively.

Overall, our implementation of the GAE algorithm for the small robot demonstrates its effectiveness in learning to navigate through an environment and reach a specific goal state. Furthermore, this algorithm can be extended to other reinforcement learning tasks in various environments, such as control and decision-making.

## ACKNOWLEDGMENT

This work was supported by the Universidad Distrital Francisco José de Caldas, specifically by the Technological Faculty. The views expressed in this paper are not necessarily endorsed by Universidad Distrital. The authors thank all the students and researchers of the research group ARMOS for their support in the development of this work.

## REFERENCES

- [1] K. P. Divakarla, A. Emadi, S. Razavi, S. Habibi, and F. Yan, "A review of autonomous vehicle technology landscape," *International Journal of Electric and Hybrid Vehicles*, vol. 11, no. 4, p. 320, 2019.
- [2] S. Jain, N. J. Ahuja, P. Srikanth, K. V. Bhadane, B. Nagaiah, A. Kumar, and C. Konstantinou, "Blockchain and autonomous vehicles: Recent advances and future directions," *IEEE Access*, vol. 9, no. 2021, pp. 130264–130328, 2021.
- [3] S. Koul and A. Eydgahi, "The impact of social influence, technophobia, and perceived safety on autonomous vehicle technology adoption," *Periodica Polytechnica Transportation Engineering*, vol. 48, no. 2, pp. 133–142, 2019.
- [4] R. B. Issa, M. Das, M. S. Rahman, M. Barua, M. K. Rhaman, K. S. N. Ripon, and M. G. R. Alam, "Double deep q-learning and faster r-CNN-based autonomous vehicle navigation and obstacle avoidance in dynamic environment," *Sensors*, vol. 21, no. 4, p. 1468, 2021.
- [5] D. Miculescu and S. Karaman, "Polling-systems-based autonomous vehicle coordination in traffic intersections with no traffic signals," *IEEE Transactions on Automatic Control*, vol. 65, no. 2, pp. 680–694, 2020.
- [6] W. Lin, X. Ren, J. Hu, Y. He, Z. Li, and M. Tong, "Fast, robust and accurate posture detection algorithm based on kalman filter and SSD for AGV," *Neurocomputing*, vol. 316, no. 2018, pp. 306–312, nov 2018.
- [7] X.-B. Jin, T.-L. Su, J.-L. Kong, Y.-T. Bai, B.-B. Miao, and C. Dou, "State-of-the-art mobile intelligence: Enabling robots to move like humans by estimating mobility with artificial intelligence," *Applied Sciences*, vol. 8, no. 3, p. 379, 2018.
- [8] H. Min, X. Wu, C. Cheng, and X. Zhao, "Kinematic and dynamic vehicle model-assisted global positioning method for autonomous vehicles with low-cost GPS/camera/in-vehicle sensors," *Sensors*, vol. 19, no. 24, p. 5430, 2019.

- [9] C. Fan, Z. Zhou, X. He, Y. Fan, L. Zhang, X. Wu, and X. Hu, "Bio-inspired multisensor navigation system based on the skylight compass and visual place recognition for unmanned aerial vehicles," *IEEE Sensors Journal*, vol. 22, no. 15, pp. 15 419–15 428, 2022.
- [10] F. Maurelli, S. Krupiński, X. Xiang, and Y. Petillot, "AUV localisation: a review of passive and active techniques," *International Journal of Intelligent Robotics and Applications*, vol. 6, no. 2, pp. 246–269, 2021.
- [11] T. Elmokadem and A. V. Savkin, "Towards fully autonomous UAVs: A survey," *Sensors*, vol. 21, no. 18, p. 6223, 2021.
- [12] X. Liu, G. V. Nardari, F. C. Ojeda, Y. Tao, A. Zhou, T. Donnelly, C. Qu, S. W. Chen, R. A. F. Romero, C. J. Taylor, and V. Kumar, "Large-scale autonomous flight with real-time semantic SLAM under dense forest canopy," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 5512–5519, 2022.
- [13] E. Pairet, J. D. Hernandez, M. Carreras, Y. Petillot, and M. Lahijanian, "Online mapping and motion planning under uncertainty for safe navigation in unknown environments," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 4, pp. 3356–3378, 2022.
- [14] I. Mir, F. Gul, S. Mir, M. A. Khan, N. Saeed, L. Abualigah, B. Abuhaija, and A. H. Gandomi, "A survey of trajectory planning techniques for autonomous systems," *Electronics*, vol. 11, no. 18, p. 2801, 2022.
- [15] Y. A. Younes and M. Barczyk, "Optimal motion planning in GPS-denied environments using nonlinear model predictive horizon," *Sensors*, vol. 21, no. 16, p. 5547, 2021.
- [16] A. Haydari and Y. Yilmaz, "Deep reinforcement learning for intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 11–32, 2022.
- [17] K.-C. Ma, L. Liu, H. K. Heidarrson, and G. S. Sukhatme, "Data-driven learning and planning for environmental sampling," *Journal of Field Robotics*, vol. 35, no. 5, pp. 643–661, 2017.
- [18] F. Martínez, F. Martínez, and E. Jacinto, "Visual identification and similarity measures used for on-line motion planning of autonomous robots in unknown environments," in *Eighth International Conference on Graphic and Image Processing (ICGIP 2016)*, 2016, pp. 321–325.
- [19] F. Martínez, C. Hernández, and A. Rendón, "A study on machine learning models for convergence time predictions in reactive navigation strategies," *Contemporary Engineering Sciences*, vol. 10, no. 25, pp. 1223–1232, 2017.
- [20] Z. Fei, Y. Wang, J. Wang, K. Liu, B. Huang, and P. Tan, "A new noise network and gradient parallelisation-based asynchronous advantage actor-critic algorithm," *IET Cyber-Systems and Robotics*, vol. 4, no. 3, pp. 175–188, 2022.
- [21] B. Peng, M. F. Keskin, B. Kulcsár, and H. Wymeersch, "Connected autonomous vehicles for improving mixed traffic efficiency in unsignalized intersections with deep reinforcement learning," *Communications in Transportation Research*, vol. 1, no. 2021, p. 100017, 2021.
- [22] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Pearson Education, 2009.
- [23] M. L. Littman, "Reinforcement learning improves behaviour from evaluative feedback," *Nature*, vol. 521, no. 7553, pp. 445–451, 2015.
- [24] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv*, no. 1707.06347, pp. 1–12, 2017.
- [25] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, P. Abbeel, and W. Zaremba, "Hindsight experience replay," *arXiv*, no. 1707.01495, pp. 1–15, 2017.
- [26] G. Li, M. Müller, V. Casser, N. Smith, D. L. Michels, and B. Ghanem, "Oil: Observational imitation learning," *arXiv*, pp. 1–21, 2018.
- [27] V. J. Hodge, R. Hawkins, and R. Alexander, "Deep reinforcement learning for drone navigation using sensor data," *Neural Computing and Applications*, vol. 33, no. 6, pp. 2015–2033, 2020.
- [28] D. Wang, T. Fan, T. Han, and J. Pan, "A two-stage reinforcement learning approach for multi-UAV collision avoidance under imperfect sensing," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3098–3105, 2020.
- [29] B. Chen, M. Xu, Z. Liu, L. Li, and D. Zhao, "Delay-aware multi-agent reinforcement learning for cooperative and competitive environments," *arXiv*, pp. 1–11, 2020.
- [30] X. Chen and P. Chaudhari, "Midas: Multi-agent interaction-aware decision-making with adaptive strategies for urban autonomous navigation," *arXiv*, pp. 1–16, 2020.
- [31] L. He, A. Nabil, and B. Song, "Explainable deep reinforcement learning for uav autonomous navigation," *arXiv*, pp. 1–12, 2020.
- [32] J. Fang, Q. Sun, Y. Chen, and Y. Tang, "Quadrotor navigation in dynamic environments with deep reinforcement learning," *Assembly Automation*, vol. 41, no. 3, pp. 254–262, 2021.
- [33] N. Yan, S. Huang, and C. Kong, "Reinforcement learning-based autonomous navigation and obstacle avoidance for USVs under partially observable conditions," *Mathematical Problems in Engineering*, vol. 2021, no. 2021, pp. 1–13, 2021.
- [34] R. B. Grando, J. C. de Jesus, V. A. Kich, A. H. Kolling, N. P. Bortoluzzi, P. M. Pinheiro, A. A. Neto, and P. L. J. Drews-Jr, "Deep reinforcement learning for mapless navigation of a hybrid aerial underwater vehicle with medium transition," *arXiv*, pp. 1–7, 2021.
- [35] M. Li, P. Sankaran, M. Kuhl, A. Ganguly, A. Kwasinski, and R. Ptucha, "Simulation analysis of a deep reinforcement learning approach for task selection by autonomous material handling vehicles," in *Winter Simulation Conference (WSC 2018)*, 2018.
- [36] F. Martínez, "Turtlebot3 robot operation for navigation applications using ros," *Tekhnê*, vol. 18, no. 2, pp. 19–24, 2021.

# A Low-Cost Wearable Autonomous System for the Protection of Bicycle Users

Daniel Mejia<sup>1</sup>, Sergio Gómez<sup>2</sup>, Fredy Martínez<sup>3</sup>  
Universidad Distrital, Francisco José de Caldas  
Bogotá D.C., Colombia

**Abstract**—A bicycle is a form of transport that not only positively impacts the health of users, and the general population by reducing pollution levels but also constitutes an accessible and affordable means of transport for developing societies. However, when coexisting with other forms of transport, the accident rate is elevated, and the risk is high. Among the factors contributing to accidents involving bicycles are collisions with motor vehicles. These accidents can occur when a motor vehicle maneuvers and does not see the bicycle or when a motorist drives distracted. These types of accidents can be avoided if cyclists and motorists are aware of the environment and respect traffic laws and safety regulations. This research aims to develop a low-cost autonomous electronic system that provides extra protection to bicycle users, particularly by making them visible to other road users on cloudy days or at night. The system uses a 32-bit processor with brightness and acceleration sensors that trigger visual alerts to both the bicycle user and possible nearby vehicles. It also monitors and logs the signals on a server for route evaluation. The laboratory successfully evaluated the prototype, demonstrating its autonomy and performance. The test results obtained demonstrate the system's capacity to provide extra protection, in addition to its robustness and accuracy.

**Keywords**—Autonomous system; bicycle users; embedded system; protection; wearable

## I. INTRODUCTION

The bicycle is a mode of transportation that has gained increasing importance in developing countries such as Colombia due to its multiple benefits [1]. First, it is an affordable and accessible form of transportation for many people who cannot afford to buy or maintain a motorized vehicle [2]. In addition, the bicycle is an environmentally friendly means of transportation, as it does not emit polluting gases or produce noise, making it an attractive option in urban environments where air pollution and noise can be severe problems [3].

Another advantage of the bicycle as a means of transportation is its positive impact on people's health [4]. Cycling as a means of transportation is a physical activity that can help improve cardiovascular fitness and burn calories, which can help prevent diseases related to sedentary lifestyles [5]. Furthermore, in developing countries where access to health services can sometimes be limited, cycling can effectively improve the population's health [6].

Cycling can also be an efficient form of transportation in environments with little or no road infrastructure [7]. In many developing countries, road infrastructure can be poor or non-existent, making cycling a practical and effective form of transportation [8]. In addition, cycling as a mode of transport can help reduce traffic congestion in cities, as it takes up less space on roads and streets.

Despite its many benefits, cycling also needs some help as a means of transportation [9]. One of the main problems is the need for adequate cycling infrastructure in many cities and rural areas [10]. This can include a need for dedicated bicycle lanes or safe and adequate parking spaces. This lack of infrastructure can make it difficult and dangerous to use bicycles as a means of transportation in some areas.

Another problem is the need for more acceptance and respect by motor vehicle drivers for bicyclists [11]. So often, drivers are not habituated to sharing the road with cyclists and may be inconsiderate or hostile towards them [12], [13]. This can make it difficult and dangerous for cyclists to share the road with motor vehicles.

Bicycle user accidents in developing countries are a serious and worrying problem [14]. Bicycle accidents can have severe consequences for those involved, including serious injury or death. In addition, bicycle accidents can have a significant economic and social impact, as they can require costly medical care and disrupt people's ability to work and support their families.

Several factors can contribute to bicycle accidents [15]. One of the main factors is the need for adequate infrastructure for bicyclists, such as dedicated bicycle lanes or safe and adequate places to park bicycles. This lack of infrastructure can make it difficult and dangerous to use bicycles as a means of transportation in some areas. However, the most critical factor is the need for more awareness and respect by motor vehicle drivers for cyclists [16]. Often, drivers are not accustomed to sharing the road with cyclists and may be inconsiderate or hostile towards them. In addition, lighting and visibility conditions on the road increase the risk of bicycle accidents.

Lack of adequate lighting on roads and streets is a significant factor in nighttime bicycle accidents [17], [18]. When there is little or no lighting, it can be difficult for bicyclists and other road users to see obstacles or potential hazards, increasing the risk of accidents [19]. In addition, poor lighting can make it difficult for motor vehicle drivers to see bicyclists, which can also increase the risk of crashes [20].

This research proposes a low-cost electronic system capable of increasing the visibility of bicycle users [21]. This system incorporates a 32-bit microcontroller with two cores capable of high-speed response and Wi-Fi and Bluetooth communication [22], [23]. This processor works hand in hand with an illumination level sensor (illuminance) to identify risk states and generate visual and acoustic alerts [24], [25].

The problem addressed in this research is the high accident rate of bicycles when coexisting with other forms of transporta-

tion, particularly motor vehicles. Bicycle accidents can occur when a motorist does not see the bicycle or when the motorist is distracted. These accidents can have severe consequences for cyclists, including injury or death.

One contributing factor to these accidents is the difficulty in seeing bicycles, especially on cloudy days or at night. Cyclists may not be visible to other road users, leading to collisions. In developing societies, where bicycles are often an affordable and accessible form of transportation, this problem is especially relevant.

To address this problem, the research aims to develop a low-cost autonomous electronic system that provides extra protection to bicycle users by making them more visible to other road users. The system uses a 32-bit processor with brightness and acceleration sensors that trigger visual alerts to both the bicycle user and nearby vehicles. It also monitors and logs the signals on a server for route evaluation. The goal is to create a robust, accurate system, and capable of providing extra protection to bicycle users to reduce the risk of accidents.

The following section presents a review of the background considered in the design of the prototype, both at the functional and implementation levels. The methods section summarizes the design criteria, its implementation, and system usage considerations. The results section presents the performance tests and the behavior recorded in laboratory conditions, and finally the conclusions section presents the summarized results.

## II. BACKGROUND

In recent years, there has been a growing interest in the development of innovative Human Activity Recognition (HAR) systems that exploit the potential of wearable devices integrated with deep learning techniques. One such example is presented by [26], who propose an innovative HAR system with the aim of recognizing the most common daily activities of a person at home. The authors use a combination of wearable sensors and deep learning techniques to accurately classify activities such as walking, sitting, and standing. They evaluate the performance of their system using a dataset of daily activities collected from 20 participants and achieve an overall accuracy of 96.7%. This study highlights the potential of using wearable devices and deep learning techniques for accurate recognition of daily activities.

Another area of research in the field of robotics and intelligent systems is the design and construction of low-cost Internet of Things (IoT) sensor meshes for remote measurement of parameters. The author in [27] present a design and construction of a low-cost IoT sensor mesh that enables the remote measurement of parameters of large-scale orchards. The authors propose the use of a mesh network of low-cost, low-power sensors to collect data on temperature, humidity, and soil moisture. The sensor mesh is designed to be easy to deploy, maintain and provides real-time data to the farmers. The system is evaluated in a commercial orchard and the results demonstrate that it can accurately measure temperature, humidity, and soil moisture.

[28] explore the technical development of a multi-user multi-modal traffic simulation platform that expands on the capabilities of traditional traffic simulators. The authors propose

a platform that combines traditional traffic simulation with virtual reality (VR) technology, allowing for multi-user and multi-modal simulations. The platform is evaluated through a case study of a complex urban intersection and the results demonstrate its ability to accurately simulate traffic flow and pedestrian behavior. This study highlights the potential of using VR technology to enhance the capabilities of traditional traffic simulation platforms and improve their accuracy in simulating complex traffic scenarios.

[29] introduce a novel low-cost solar-powered wearable assistive technology (AT) device, whose aim is to provide continuous, real-time object recognition to ease the finding of the objects for visually impaired (VI) people in daily life. The authors propose the use of a low-cost, solar-powered wearable device equipped with a camera and a deep learning-based object recognition system. The device is evaluated through user studies with VI individuals and the results demonstrate its ability to accurately recognize objects in real-time and improve the independence of VI individuals in their daily life.

[30] present a new “active mask” paradigm, in which the wearable device is equipped with smart sensors and actuators to both detect the presence of airborne pathogens in real time and take appropriate action to mitigate the threat. The authors propose the use of an active mask equipped with sensors for detecting airborne pathogens and actuators for releasing disinfectants. The device is evaluated through laboratory tests and the results demonstrate its ability to accurately detect pathogens and effectively release disinfectants.

[31] aim at helping the mobility-challenged individuals with a novel robotic companion, which is a walker-type mobile robot capable of accompanying the human user and keeping user at the center for protection and possible power assistance. The authors propose a robot that can assist mobility-challenged individuals with navigation, object recognition, and power assistance. The robot is evaluated through user studies with mobility-challenged individuals and the results demonstrate its ability to improve the independence and safety of the users.

[32] use VR 360° panoramic technology to develop a virtual wetland ecological system for applications in environmental education. The authors propose the use of VR technology to create an immersive virtual wetland ecosystem for educational purposes. The virtual wetland is evaluated through a user study with students and the results demonstrate its ability to improve the understanding and engagement of students in learning about wetland ecology. This study highlights the potential of using VR technology in environmental education to create immersive and interactive learning experiences.

[33] present a new “active mask” paradigm, similar to the one presented by [30], in which the wearable device is equipped with smart sensors and actuators to both detect the presence of airborne pathogens in real time and take appropriate action to mitigate the threat. The authors propose an active mask equipped with sensors for detecting COVID-19 and actuators for releasing disinfectants. The device is evaluated through laboratory tests and the results demonstrate its ability to accurately detect COVID-19 and effectively release disinfectants.

[34] introduce an artificial intelligence-powered multi-modal robotic sensing system (M-Bot) with an all-printed

mass-producible soft electronic skin-based human-machine interface. The authors propose a robot that can sense and interact with its environment through an electronic skin interface. The robot is equipped with AI algorithms for object recognition and tactile sensing. The robot is evaluated through laboratory tests and the results demonstrate its ability to accurately sense and interact with its environment through its electronic skin interface.

[35] develop an air quality monitoring platform that comprises a wearable device embedding low-cost metal oxide semiconductor (MOS) gas sensors, a PM sensor, and a smartphone for collecting the data using Bluetooth Low Energy (BLE) communication. The authors propose a wearable device for monitoring air quality, specifically for PM and gas pollutants. The device is evaluated through laboratory tests and the results demonstrate its ability to accurately measure PM and gas pollutants in the air.

### III. METHODS

The project's objective was to develop a low-cost wearable system (accessible to the average user) with high processing and communication capabilities. To meet these criteria, a vest was developed with embedded electronics equipped with sensors and display units, both for the user and individuals on the road. The control unit selected was the ESP32 SoC (System On a Chip) from Espressif Systems, which is supported by a 32-bit, dual-core Tensilica Xtensa LX6 microprocessor. This processor supports a clock frequency of up to 240 MHz, with a performance of up to 600 DMIPS (Dhrystone Million Instructions Per Second). In addition, the SoC has Wi-Fi/Bluetooth functionalities that cover the protection system's communication requirements.

The protective vest also incorporates other essential safety elements for urban cyclists, among them are the following:

- **Reflective elements:** The vest includes reflective stripes and other reflective elements to help make the wearer more visible to other road users, especially in low-light conditions.
- **High visibility colors:** The design of the vest uses brightly colored and evident materials, such as neon yellow and orange, to help improve visibility.
- **LED lights:** A signaling system with LED lights have been incorporated into the vest to increase visibility. The protection system control unit controls these lights.
- **Breathability:** Breathable materials have been used in the vest design, which is especially important if you are cycling in hot climates to avoid overheating.
- **Fit:** The vest incorporates adjustable straps to fit different body types. To improve safety, a good fit should be ensured, and comfort should be improved.
- **Durability:** The vest must be made of durable materials that can withstand regular use and wear.
- **Water resistance:** Special care was taken to ensure that the vest would function in wet conditions. In

addition, a water-resistant vest can help keep the wearer dry and comfortable.

A series of sensors were incorporated into the control unit, including the illuminance sensor (luminous flux per unit area). The BH1750 sensor was used, a digital light sensor that measures in Lux (lumen/m<sup>2</sup>). The sensor uses I<sup>2</sup>C communication, has high precision, and ranges from 1 to 65535 lx. This sensor was calibrated in the laboratory with a reference lux meter in a dark room with lighting control (Fig. 1). Low values correspond to low illumination conditions, and high values to high illumination conditions. The measured error averaged 4.8%, with maximum values of 14.6%.

Other sensors and peripherals incorporated into the system include:

- A U-Blox NEO-6M GPS (Global Positioning System) U-Blox NEO-6M, a high-accuracy, low-cost GPS module for global tracking.
- A GY-521 MPU-6050 module incorporating a gyroscope and accelerometer, both with MEMS technology, has six axes of freedom (three axes of the accelerometer and three axes of the gyroscope).
- Two 32×8 pixel LEDs grid with Bluetooth to alert people close to the cyclist.
- One 84×48 pixel monochrome graphic display with PCD8544 controller for communication with the user.
- A small Lithium-ion (Li-ion) battery pack.

A road safety vest for cyclists is a portable system designed to increase the visibility and safety of cyclists when riding on the road. Functionally, the essential part is the control unit, based in our case on the ESP32 SoC, which is responsible for processing the inputs coming from the sensors and controlling the output of the vest. The following describes in detail how the control unit of the road safety vest works with illuminance sensors, gyroscope, accelerometer, GPS, and LED displays to show the status to vehicle drivers and a display to inform the user.

First, let's consider the role of the illuminance sensors. These sensors detect the ambient light level in the environment where the vest is being worn. This information is vital because the vest's LED displays, which communicate the wearer's status to vehicle drivers, need to be bright enough to be seen in low-light conditions but not too bright to be distracting in bright-light conditions. The SoC control unit will use the illuminance sensors' data to adjust the LED displays' brightness accordingly. The information displayed is congruent with the state of the cyclist, which is supported by the accelerometer. Therefore, this information is accentuated in the LED array, as well as if he/she is stopped or brakes suddenly.

Next, let's consider the role of the gyroscope and accelerometer. These sensors are used to detect the movement and orientation of the vest. The gyroscope measures angular velocity, which allows the microcontroller control unit to determine the vest's orientation in space. The accelerometer measures acceleration, which allows the microcontroller control unit to determine the vest's movement. This information is helpful for several reasons. For example, suppose the vest is

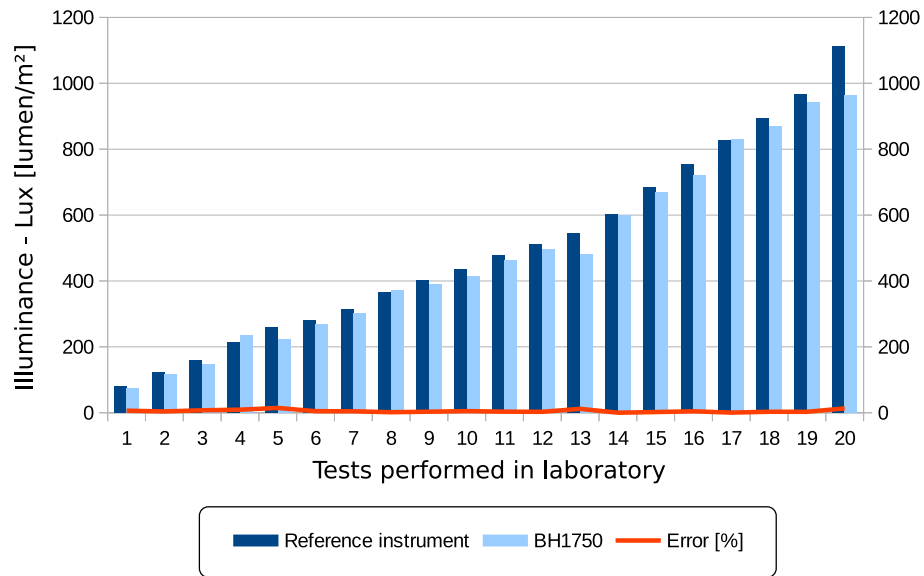


Fig. 1. Illuminance sensor vs. reference sensor characteristics

worn while cycling. In that case, the accelerometer can detect when the wearer is braking or accelerating, which can activate the LED displays to alert vehicle drivers of the wearer's intentions. Likewise, the gyroscope can detect when the wearer is turning, which can also activate the LED displays to alert vehicle drivers of the wearer's intentions.

The GPS (Global Positioning System) sensor determines the vest's location. This information is used by the SoC control unit to display the wearer's location on the display screen and to report the tracking information to a web server, which can be helpful for navigation purposes. Additionally, the GPS data can activate the LED displays when the vest is approaching a location where it is vital for the wearer to be more visible, such as a busy intersection.

The LED displays on the vest are used to communicate the wearer's status to vehicle drivers. For example, the LED displays can be activated to show a solid red light when the wearer is braking, a solid yellow light when the wearer is turning, and a solid green light when the wearer is accelerating. The LED displays can also display a flashing red light when the vest is approaching a location where it is crucial for the wearer to be more visible, such as a busy intersection.

Finally, the vest should also have a display screen to inform the user of various information, such as the vest's battery level, the current time, and the wearer's location on a map. The SoC control unit can use the data from the various sensors to update this information on the display screen in real time.

The SoC control unit of the road safety vest incorporates illuminance sensors, a gyroscope, accelerometer, GPS, LED displays to show the status to vehicle drivers, and a display to inform the user should function by using the data from the various sensors to control the output of the vest. The illuminance sensors are used to adjust the brightness of the LED displays, the gyroscope and accelerometer are used to detect the vest's movement and orientation, the GPS sensor is used to determine the vest's location, and the LED displays

are used to communicate the condition of the cyclist to other road users, primary drivers of motorized vehicles (Fig. 2).

#### IV. RESULTS AND DISCUSSION

The performance evaluation of the system was aimed at ensuring that the vest effectively increases cyclists' visibility and safety when riding on the road. In this regard, firstly, the role of illuminance sensors was considered. The performance of these sensors can be evaluated by measuring the accuracy of the ambient light level readings they provide. This was done by comparing the sensor readings with the actual ambient light level in the environment using a light meter. The sensors must provide accurate readings because the vest's LED displays, which are used to communicate the wearer's status to vehicle drivers, must be bright enough to be seen in low light conditions but not too bright to be distracting in bright light conditions (Fig. 3).

Another critical part of the testing considered the role of the gyroscope and accelerometer. The performance of these sensors was evaluated by measuring the accuracy of the angular velocity and acceleration readings they provide. This was done by comparing the sensor readings with the actual angular velocity and acceleration of the vest using a separate device (a gyroscope and a high-precision accelerometer). The sensors must provide accurate readings because they are used to detect the vest's motion and orientation, which activate the LED displays that alert vehicle drivers of the wearer's intentions.

The performance of the GPS (Global Positioning System) sensor was evaluated by measuring the accuracy of the location readings it provides. This was done by comparing the sensor readings with the actual location of the vest using an independent GPS device and a map. Accurate location readings are essential for the sensor because the vest's display uses this information to show the user's location. This information is also used to track the user on the Internet. In addition, the



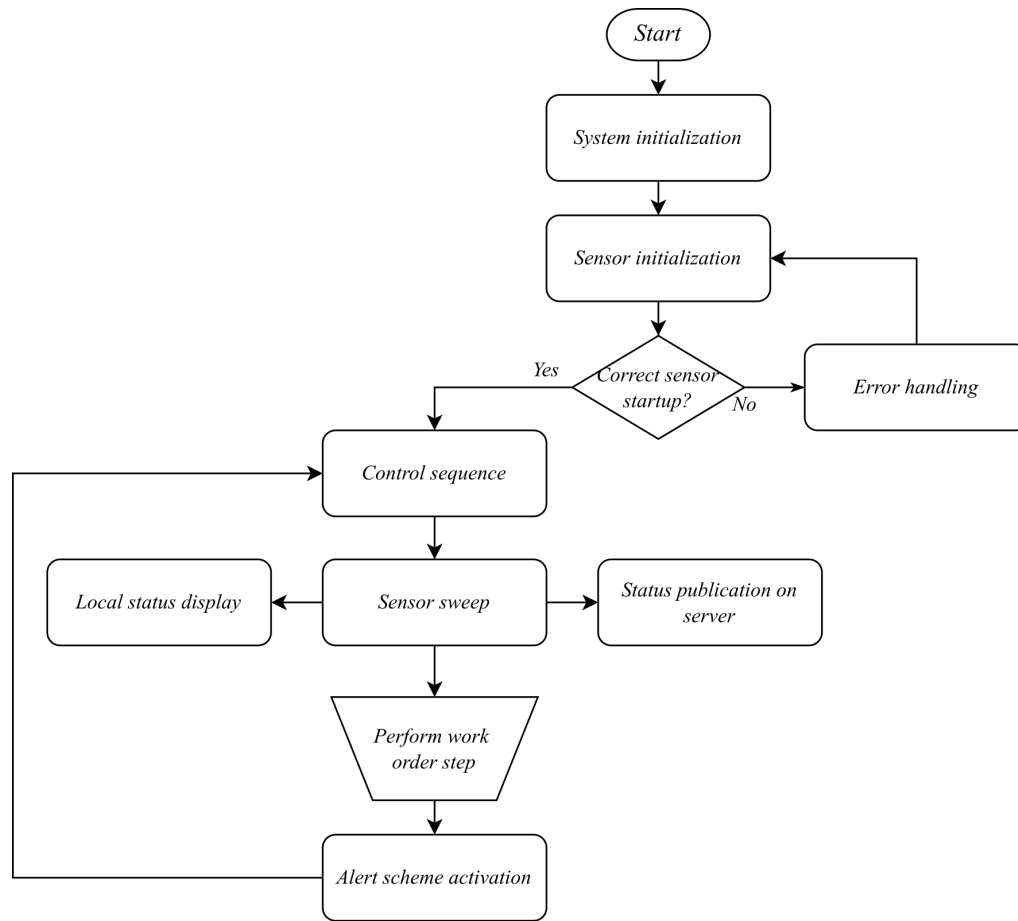


Fig. 2. Proposed operating strategy

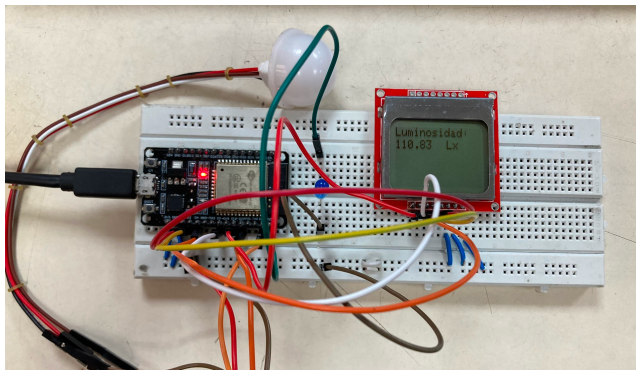


Fig. 3. Illuminance sensor tests

at different angles as they communicate the wearer's status to vehicle drivers.

Finally, the performance of the display screen that informs the user of various data, such as the vest's battery level, the current time and the user's location coordinates, was evaluated by measuring the accuracy and readability of the information it displays. This was done by comparing the information displayed on the screen with that of another device and by viewing the screen from different positions while wearing the vest. The display screen must provide accurate and legible information because it is used to inform the wearer of various vital details. In addition, the logging of information of all variables detected by the vest to the web server was verified, which should include the date and time information, and match the information displayed to the user in real-time (Fig. 4).

GPS data is used to activate the LED displays when the vest approaches a location where it is vital for the wearer to be more visible, such as a busy intersection.

The performance of the LED displays was evaluated by measuring their visibility from different distances and angles. This was done by observing the displays from different positions while wearing the vest and comparing the visibility to a standard. The displays must be evident from a distance and

In summary, the cycling safety vest's performance meets the design profile's performance expectations. It has an excellent response to changes in illuminance and provides high visibility by combining LED lights with bright, high-visibility colors. The response characteristics to sudden changes in acceleration and geographic location increase user protection in situations beyond their control. The system will be further tested for mechanical robustness, ergonomics and comfort to meet additional design parameters.



Fig. 4. Tracking of system variables

## V. CONCLUSION

In conclusion, the proposed safety vest for cyclists is a portable and innovative device that aims to improve the visibility and safety of cyclists on the road. The vest incorporates illuminance sensors, a gyroscope, an accelerometer, GPS, and LED displays to show the status to vehicle drivers and a display to inform the user. The SoC that controls the vest is responsible for processing the information from the different sensors and controlling the output of the vest. The illuminance sensors measure the ambient light level in the environment in which the vest is worn, which is vital for adjusting the brightness of the LED displays accordingly. The gyroscope and accelerometer, meanwhile, measure the movement and orientation of the vest and are used to activate the LED displays and alert vehicle drivers of the wearer's intentions. The GPS sensor determines the location of the vest. It is used to display the wearer's location on the screen and to activate the LED displays when the vest approaches a location where the wearer needs to be more visible, such as a busy intersection. All detected information is transmitted in real-time to a web server for logging and tracking. LED displays are used to communicate the user's status to vehicle drivers, showing solid red light when the user brakes, solid yellow light when turning, and a solid green light when accelerating. They can also be used to display a flashing red light when the vest approaches a location where the wearer needs to be more visible. Finally, the vest also has a display that informs the user of the vest's battery level, the current time and the user's location on a map.

The results of the laboratory tests show that the vest is a well-designed and effective device that can significantly improve the visibility and safety of cyclists. The ability to adjust the brightness of the LED displays based on the ambient light level detected by the illuminance sensors ensures that the displays are always visible to vehicle drivers, regardless of lighting conditions. In addition, the gyroscope and accelerometer allow the vest to detect the wearer's movement and orientation, which can activate the LED displays and alert drivers of the wearer's intentions. Finally, the GPS sensor allows the vest to determine its location, which can be displayed on a map on the display screen for navigation purposes and used to

activate the LED displays when the vest approaches a location where the wearer must be more visible.

Future work on this vest could include the integration of additional sensors, such as a heart rate monitor, to provide the user with even more information about their physical condition while cycling. Another potential improvement could be the integration of a communication system, such as a two-way radio, to allow the user to communicate with other cyclists or vehicle drivers in their vicinity. Additionally, the vest could be integrated with a smart phone app that allows the user to remotely control the vest's functions and view real-time data from the sensors. Finally, further testing and evaluation of the vest in real-world conditions, including in different weather conditions, would help to further optimize its performance and effectiveness.

## ACKNOWLEDGMENT

This work was supported by the Universidad Distrital Francisco José de Caldas, specifically by the Technological Faculty. The views expressed in this paper are not necessarily endorsed by Universidad Distrital. The authors thank all the students and researchers of the research group ARMOS for their support in the development of this work.

## REFERENCES

- [1] P. K. Machavarapu and S. Ram, "Review on public bike share schemes in large developing cities: A case study of delhi, india," *Case Studies on Transport Policy*, vol. 10, no. 4, pp. 2075–2091, 2022.
- [2] L. Ordinez, C. Buckle, S. A. Kaminker, D. Firmenich, D. Barry, and A. Aguirre, "Assessing cycling social feasibility in a medium-size patagonian city," *Transportation Research Part D: Transport and Environment*, vol. 92, no. 2021, p. 102720, 2021.
- [3] R. Sabyrbekov and I. Overland, "Why choose to cycle in a low-income country?" *Sustainability*, vol. 12, no. 18, p. 7775, 2020.
- [4] R. Gore, C. J. Lynch, C. A. Jordan, A. Collins, R. M. Robinson, G. Fuller, P. Ames, P. Keerthi, and Y. Kandukuri, "Estimating the health effects of adding bicycle and pedestrian paths at the census tract level: Multiple model comparison," *JMIR Public Health and Surveillance*, vol. 8, no. 8, p. e37379, 2022.
- [5] J. Raustorp and T. Koglin, "The potential for active commuting by bicycle and its possible effects on public health," *Journal of Transport & Health*, vol. 13, no. 2019, pp. 72–77, 2019.

- [6] D. G. T. Whitehurst, D. N. DeVries, D. Fuller, and M. Winters, "An economic analysis of the health-related benefits associated with bicycle infrastructure investment in three canadian cities," *PLOS ONE*, vol. 16, no. 2, p. e0246419, 2021.
- [7] L. B. Meuleners, M. Stevenson, M. Fraser, J. Oxley, G. Rose, and M. Johnson, "Safer cycling and the urban road environment: A case control study," *Accident Analysis & Prevention*, vol. 129, no. 2019, pp. 342–349, 2019.
- [8] L. Meng, "Political economy and cycling infrastructure investment," *Transportation Research Interdisciplinary Perspectives*, vol. 14, no. 2022, p. 100618, 2022.
- [9] L. Meng, S. Somenahalli, and S. Berry, "Policy implementation of multi-modal (shared) mobility: review of a supply-demand value proposition canvas," *Transport Reviews*, vol. 40, no. 5, pp. 670–684, 2020.
- [10] N. H. Tran, S.-H. Yang, and T. Huang, "Comparative analysis of traffic-and-transportation-planning-related indicators in sustainable transportation infrastructure rating systems," *International Journal of Sustainable Transportation*, vol. 15, no. 3, pp. 203–216, 2020.
- [11] A. K. Debnath, N. Haworth, A. Schramm, K. C. Heesch, and K. Somoray, "Factors influencing noncompliance with bicycle passing distance laws," *Accident Analysis & Prevention*, vol. 115, no. 2018, pp. 137–142, 2018.
- [12] T. Uijtdeuwilgen, M. B. Ulak, G. J. Wijlhuizen, F. Bijleveld, A. Dijkstra, and K. T. Geurs, "How does hourly variation in exposure to cyclists and motorised vehicles affect cyclist safety? a case study from a dutch cycling capital," *Safety Science*, vol. 152, no. 2022, p. 105740, 2022.
- [13] E. Rubie, N. Haworth, D. Twisk, and N. Yamamoto, "Influences on lateral passing distance when motor vehicles overtake bicycles: a systematic literature review," *Transport Reviews*, vol. 40, no. 6, pp. 754–773, 2020.
- [14] P. M. Salmon, M. Naughton, A. Hulme, and S. McLean, "Bicycle crash contributory factors: A systematic review," *Safety Science*, vol. 145, no. 2022, p. 105511, 2022.
- [15] S. A. Useche, S. O'Hern, A. Gonzalez-Marin, J. Gene-Morales, F. Alonso, and A. N. Stephens, "Unsafety on two wheels, or social prejudice? proxying behavioral reports on bicycle and e-scooter riding safety – a mixed-methods study," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 89, no. 2022, pp. 168–182, 2022.
- [16] S. Cordovez, E. Ortiz-Prado, E. Vasconez, F. Andrade, K. Simbaña-Rivera, L. Gómez-Barreno, and R. C. McIlroy, "Bicycling-related mortality in ecuador: A nationwide population-based analysis from 2004 to 2017," *Sustainability*, vol. 13, no. 11, p. 5906, 2021.
- [17] R. Abdur, K. Aya, K. Teppei, and K. Hisashi, "A mechanism to enhance bicycle conspicuity and visibility and increase detection distances: New insights into bicycle safety," *IATSS Research*, vol. 45, no. 2, pp. 241–250, 2021.
- [18] M. Asgarzadeh, D. Fischer, S. K. Verma, T. K. Courtney, and D. C. Christiani, "The impact of weather, road surface, time-of-day, and light conditions on severity of bicycle-motor vehicle crash injuries," *American Journal of Industrial Medicine*, vol. 61, no. 7, pp. 556–565, 2018.
- [19] K. Billah, H. O. Sharif, and S. Dessouky, "Analysis of bicycle-motor vehicle crashes in san antonio, texas," *International Journal of Environmental Research and Public Health*, vol. 18, no. 17, p. 9220, 2021.
- [20] A. N. Cihan and G. N. Gugul, "An indoor smart lamp for environments illuminated day time," sep 2020.
- [21] A. Suryana, F. P. Lismana, R. M. Rachmat, S. D. Putra, and M. Artiyasa, "Implementation of weather station for the weather reality in a room," in *6th International Conference on Computing Engineering and Design (ICCED 2020)*, no. 9415799. IEEE, oct 2020.
- [22] I. Desnanjaya, A. Gede, I. Aditya, I. Komang, and M. Sumaharja, "Room monitoring uses ESP-12e based DHT22 and BH1750 sensors," *Journal of Robotics and Control (JRC)*, vol. 3, no. 2, pp. 205–211, feb 2022.
- [23] M. Ahmad, S. Z. M. Noor, N. F. A. Rahman, and F. A. Haris, "Lux meter integrated with internet of things (IoT) and data storage (LMX20)," mar 2021.
- [24] J. Gao, J. Luo, A. Xu, and J. Yu, "Light intensity intelligent control system research and design based on automobile sun visor of BH1750," no. 7979192, pp. 3957 – 3960, 2017.
- [25] A. Pan and N. Wang, "Design and implementation of crop automatic diagnosis and treatment system based on internet of things," *Journal of Physics: Conference Series*, vol. 1883, no. 1, p. 012062, apr 2021.
- [26] V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M. Mordonini, and I. D. Munari, "IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8553–8562, 2019.
- [27] L. Varandas, J. Faria, P. Gaspar, and M. Aguiar, "Low-cost IoT remote sensor mesh for large-scale orchard monitoring," *Journal of Sensor and Actuator Networks*, vol. 9, no. 3, p. 44, 2020.
- [28] J. Miller, V. Kalivarapu, M. Holm, T. Finseth, J. Williams, and E. Winer, "A flexible multi-modal multi-user traffic simulation for studying complex road design," in *Volume 9: 40th Computers and Information in Engineering Conference (CIE)*. American Society of Mechanical Engineers, 2020.
- [29] B. Calabrese, R. Velázquez, C. Del-Valle-Soto, R. de Fazio, N. I. Giannoccaro, and P. Visconti, "Solar-powered deep learning-based recognition system of daily used objects and human faces for assistance of the visually impaired," *Energies*, vol. 13, no. 22, p. 6104, 2020.
- [30] N. V. R. Masna, R. R. Kalavakonda, R. Dizon, A. Bhuniaroy, S. Mandal, and S. Bhunia, "The smart mask: Active closed-loop protection against airborne pathogens," *arXiv*, pp. 1–8, 2020.
- [31] T. Shen, M. R. Afsar, H. Zhang, C. Ye, and X. Shen, "A 3d computer vision-guided robotic companion for non-contact human assistance and rehabilitation," *Journal of Intelligent and Robotic Systems*, vol. 100, no. 3-4, pp. 911–923, 2020.
- [32] K.-L. Ou, S.-T. Chu, and W. Tarnq, "Development of a virtual wetland ecological system using VR 360° panoramic technology for environmental education," *Land*, vol. 10, no. 8, p. 829, 2021.
- [33] R. R. Kalavakonda, N. V. R. Masna, A. Bhuniaroy, S. Mandal, and S. Bhunia, "A smart mask for active defense against coronaviruses and other airborne pathogens," *IEEE Consumer Electronics Magazine*, vol. 10, no. 2, pp. 72–79, 2021.
- [34] Y. Yu, J. Li, S. A. Solomon, J. Min, J. Tu, W. Guo, C. Xu, Y. Song, and W. Gao, "All-printed soft human-machine interface for robotic physicochemical sensing," *Science Robotics*, vol. 7, no. 67, pp. 1–6, 2022.
- [35] S. Palomeque-Mangut, F. Meléndez, J. Gómez-Suárez, S. Frutos-Puerto, P. Arroyo, E. Pinilla-Gil, and J. Lozano, "Wearable system for outdoor air quality monitoring in a WSN with cloud computing: Design, validation and deployment," *Chemosphere*, vol. 307, no. 2022, p. 135948, 2022.

# An Automated Impact Analysis Approach for Test Cases based on Changes of Use Case based Requirement Specifications

Adisak Intana<sup>1</sup>, Kanjana Laosen<sup>2</sup>, Thiwatip Sriraksa<sup>3</sup>  
College of Computing, Prince of Songkla University, Phuket, Thailand

**Abstract**—Change Impact Analysis (CIA) is essential to the software development process that identifies the potential effects of changes during the development process. The changing of requirements always impacts on the software testing because some parts of the existing test cases may not be used to test the software. This affects new test cases to be entirely generated from the changed version of software requirements specification that causes a considerable amount of time and effort to generate new test cases to re-test the modified system. Therefore, this paper proposes a novel automatic impact analysis approach of test cases based on changes of use case based requirement specification. This approach enables a framework and CIA algorithm where the impact of test cases is analysed when the requirement specification is changed. To detect the change, two versions as before-change and after-change of the use case model are compared. Consequently, the patterns representing the cause of variable changes are classified and analysed. This results in the existing test cases to be analysed whether they are completely reused, partly updated as well as additionally generated. The new test cases are generated automatically by using the Combination of Equivalence and Classification Tree Method (CCTM). This benefits the level of testing coverage with a minimised number of test cases to be enabled and the redundant test cases to be eliminated. The automation of this approach is demonstrated with the developed prototype tool. The validation and evaluation result with two real case studies from Hospital Information System (HIS) together with perspective views of practical specialists confirms the contribution of this tool that we seek.

**Keywords**—Change impact analysis approach; test case; black-box testing; use case based requirement specification; combination of equivalence and classification tree method

## I. INTRODUCTION

Software testing is one of the most necessary and integral stages in the software development life cycle [1]. It consists of a set of activities that are performed by using a systematic approach. This enables a System Under Test (SUT) to be executed on a set of test cases to ensure that the software system is free from error. Black-box testing is one of the software testing methods that examines the functionality of software without knowing the internal structures or mechanisms. It is sometimes called specification-based testing as this type of software testing is being performed to validate the complete or integrated software based on the Software Requirements Specification (SRS) document. Test cases are usually derived from software artifacts such as specifications and design. Software testers initially go through the requirement document to understand requirements and specifications before designing and generating test cases.

Recently, black-box testing techniques have been widely applied in many practical software development domains including Point of Sales (POSs) [2], Internet of Things (IoTs) [3], School Payment Information System [4] and e-Learning system [5]. The result of this application can confirm that black-box testing can guarantee the correctness of the system functionality whether it meets the expected users' needs. Furthermore, in the software testing community, several black-box testing techniques have been introduced. The widely known black-box testing techniques include Equivalence Class Partition (ECP) [1], Boundary Value Analysis (BVA) [1], Classification Tree Method (CTM) [6], [7], Decision Table-Based Testing [1], State Transition Testing [1]. These influence the benefit which maximises test case coverage with a minimised number of test cases. However, one of the research challenges in software testing is that since the generated test case is normally derived from the SRS document, changing requirements always impacts on the existing previous test cases which are unable to be retested in the modified system [8], [9]. As the advanced technology demands businesses grow and evolve their needs, the requirement of systems becomes more complex and continuously changes. Over time, some domain features may not be needed, therefore either removing or replacing them with a new feature, whilst may be refined [8], [9]. Whenever the software requirement is changed, the testing process is affected. This leads to new test cases being created from the changed version of software requirements specification. Consequently, this causes considerable time and resource effort to generate test cases to test all new systems to ensure that all components in the system are connected and functioning correctly. Thus, Change Impact Analysis (CIA) for testing is needed to determine how much software testing and test automation should be performed by software testers after the change.

Therefore, in our previous work [10], we proposed the conceptual vision of impact analysis framework of test cases based on changes of use case based requirements. This framework analyses the impact on changes of test cases when variables in use case specification are changed. Five patterns of atomic changes were designed specifically according to the change of use case specification affecting the test case as 1) change of variable name, 2) change of variable type, 3) change of variable range value, 4) change of number of variables, and 5) change of variable order. The result of this impact analysis enables the existing test cases to be completely reused, partly updated as well as additionally generated. The modified version of test cases is generated with the Combination of Equivalence

and Classification Tree Method (CCTM), the hybrid testing generation approach of ECP with CTM testing technique. The combination of testing technique increases the level of testing coverage and reduces the redundant test cases. The effectiveness of this framework was demonstrated manually with the real case study, *Kidney Failure Diagnosis (KFD) system*.

The work proposed in this paper is extended from our previous work [10]. This paper presents an automated impact analysis approach of test cases based on changes of use case based requirement specification. The automation of this approach is demonstrated by an automatic prototype tool in which the CIA algorithm of the approach is implemented in this tool. To detect changes, the tool provides consistency checking by comparing two versions as before-change and after-change of use case XML files. Five patterns of atomic changes designed specifically are encoded into this tool to analyse and classify the cause of the change. The tool also performs the analysis of level of change impact on the before-change version of test cases that are classified into four groups as *no-change*, *delete*, *update* and *create* categories. To create the new test cases, CCTM was implemented in the tool. Furthermore, we show the applicability and efficiency of the proposed automatic approach and tool with two case studies formulated from real-world Hospital Information System (HIS) with perspective views of practical specialists. The result of this test case generation from these case studies by the tool was compared with the expected test cases that were calculated manually to validate the accuracy and effectiveness of both approach and implementation. The reusability level of test cases was also evaluated by the case studies. Moreover, the satisfactory level of the proposed approach and tool was evaluated from practical specialists for potential application in the future.

The rest of the paper is structured as follows. Section II gives the necessary background and related work. Then, Section III describes and demonstrates the proposed impact analysis approach and the detailed algorithm with our experienced case study. In Section IV, we demonstrate the proof of concept of our proposed approach through the evaluation of implemented prototype tool for impact analysis of test cases with two real-world case studies. Section V presents lesson learned and discussions of this study. Finally, conclusions and areas of future work are summarised in Section VI.

## II. RELATED WORK

There are some research studies dealing with the application of the CIA approach to analyse the effect level of test case changes [11], [12], [9], [8], [13]. Some of these focus on the impact analysis of test cases from changes in database schema such as the work proposed by [12], [13]. The author in [12] presented a tool for analysing the changes of database schema encoding in terms of embedded SQL in the source code. The database schema change is detected from the recorded log file before the impact of the corresponding embedded SQL in the source code is analysed and new test cases are finally generated. In [13], the impact of changes in database schema requirements such as addition, deletion, modification or change of some schema values is analysed before a set of test cases corresponding to the change is generated.

Some research studies proposed a CIA framework for test cases from the change of specification [11], [9], [8], [14]. The author in [11], [9] proposed the impact analysis framework of test cases from the change of use case specification. The author in [11] proposed a test case selection framework based on the change of two versions of use case description. These two versions of the use case description are encoded in XML file format and used as input for this framework. Requirement validation matrix mapping the use case with its corresponding generated test cases is used to check and track the change of use cases. When there are changes in the use case, the existing test case affected from the change is considered whether it can be reusable. The author in [9] extended the capability of [11] in which the requirement validation matrix is generated automatically from the old version of use case description when it does not exist. However, both former research studies only analyse the impact results in either the existing test case from the previous version can be reused or the new test case must be generated. They do not classify the cause of change into a pattern which may help testers to analyse the change of test cases faster and more efficiently. Furthermore, their framework does not support the automatic test case generation in the case that new test cases are required to be generated due to the added feature or requirement. Recently, [14] proposed a CIA approach for test cases affected from the change of inputs or outputs of functional requirements. The approach offers the change process enabling the change control over functional requirements, test cases and database schema. The rollback mechanism is also provided to support in the case of cancelling of changes. Similar to other earlier mentioned studies, the cause of change is not classified and the level of reusability is not specified in this work.

The closely related work is proposed by [8]. They proposed the impact analysis framework of test cases from the change of use case specification. They introduced seven patterns indicating the cause of changes into their impact analysis tool. This includes the change of 1) variable name, 2) data type, 3) variable value, 4) variable tag, 5) order, 6) link and 7) variable number. They are used to analyse the two versions of HTML document and XML schema file representing the previous and changed version of input and output specified in the Graphic User Interface (GUI) of web application. ECP and BVA testing techniques are used to generate the new test case. From our literature reviews, we can conclude that there are few attempts at the specification level. Therefore, our challenge is to contribute an impact analysis approach for test case based on changes of requirement specification.

## III. MATERIALS AND APPROACH

### A. Framework of Impact Analysis of Test Cases

A conceptual overview of an impact analysis framework of test cases based on changes of use case proposed in our preliminary work [10] as shown in Fig. 1. This framework is divided into four steps: 1) variables in two versions as before-change and after-change of the use case specification model are extracted. Then, 2) the extracted variables from these two versions are compared. This results in the variables affected from the requirement change to be identified. The cause of changes is also identified. There are five patterns of atomic changes classified for the cause of changes that are a) change of

variable name, b) change of variable type, c) change of variable range value, d) change of number of variables and e) changes of variable order. After that 3) the impact analysis of test cases is performed. This enables the corresponding test case from the before-change version recorded on the database to be retrieved in order to analyse the level of change impact. Consequently, the analysed test cases are classified into four groups that are *no-change*, *delete*, *update* and *create* categories. In the case of no change effect, the test case is totally reused in the new version. However, if change impacts of test cases are detected, either the affected test case is updated giving it improvement or 4) the new test case is fully generated according to the additional requirement by using CCTM technique.

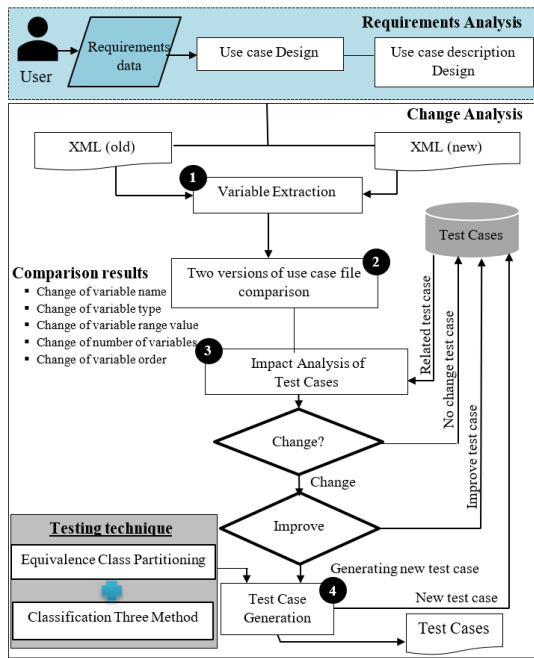


Fig. 1. Framework of impact analysis of test cases [10]

### B. Case Study

To demonstrate the effectiveness of our approach, we formulate our case study, Medical Record System (MRS) from a real-world system, Hospital Information System (HIS)<sup>1</sup>, an open-source software based on Java platform. The system was deployed over 90 hospitals in Thailand. This system has been chosen as it is a safety-critical system related to human diagnosis that requires a high degree of correctness. Furthermore, we select two real case studies from the subsystem that has the requirements changing issues, these types of modules require a high degree of accuracy calculation and measurement. To explore the effectiveness of our approach, we chose Kidney Failure Diagnosis (KFD) and Online Nursing Assessment (ONA) subsystems from HIS. Table I shows the criteria of the selected case studies from the different level of complexity. Number of implemented use cases is the first criteria to demonstrate the complexity of system. As the Cartesian product method is used in testing technique in which all variable range value are multiplied together to generate test

cases, number of variable is the second criteria. Number of change requests influencing five patterns of atomic changes is the other criteria used for case study selection. As can be seen in this table, KFD is a simple case study that is used to demonstrate and evaluate how the proposed approach and its implemented prototype tool work. ONA is a more complex case study compared to KFD in terms of number of use cases, variables and change requests. Moreover, it contains the change requests covering all patterns of atomic changes. Therefore, ONA is more focused on the evaluation of the efficiency of the approach and implementation. Specifications of these systems are described in the following sections.

TABLE I. CASE STUDIES CHARACTERISTICS

System	# use cases	# variables	# change requests	# change of				
				variable name	variable type	variable range value	number of variables	variable order
KFD	3	6	3	×	×	-	×	×
ONA	5	19	4	×	×	×	×	×

1) *Kidney Failure Diagnosis (KFD) Subsystem*: KFD is a subsystem in MRS that records the history of patient treatment. KFD subsystem provides the result of kidney failure analysis to physicians helping them to diagnose whether their patients kidney disorders are abnormal. The Chronic Kidney Disease (CKD) of a patient is calculated and used to interpret the stage of kidney disease by this system. This calculation is based on two factors: *Glomerular Filtration Rate (GFR)* and *Uric Acid Creatinine (UO)*. *GFR* is a calculation that is used to determine how well the kidneys function by estimating the blood that is filtered by the kidneys. *GFR* is normally calculated by using mathematical formula including the *Age*, *Gender* and *Serum Creatinine (Scr)* of patients. Failure of the patient kidney is interpreted by determining *GFR* values together with the effects of *UO*. This results in the stage of kidney failure to be interpreted and suggests what treatment is applied to the patient. After this diagnosis and applying the treatment to the patient, the nurse records the information regarding the measured factors, interpretation and diagnosis results into the system, so that the physician can view the history of treatment in order to monitor the patient disease later.

**Requirements.** The requirements of this system are summarised as follows:

- **REQ1:** The system shall automatically calculate *GFR* value for the patient from *Gender*, *Age* and *Scr*.
- **REQ2:** The system shall accurately perform the kidney failure state from *GFR* and *UO* value mapping in five stages.
  - *Stage 1 (Risk Stage):*  $GFR \geq 90$  AND  $UO > 300$  mg/g
  - *Stage 2 (Injury):*  $GFR = 60-89$  AND  $UO > 300$  mg/g
  - *Stage 3 (Failure):*  $GFR = 30-59$  AND ( $UO = 30-300$  mg/g OR  $UO > 300$  mg/g)
  - *Stage 4 (Loss):*  $GFR = 15-29$  AND ( $UO <$

<sup>1</sup><http://www.opensource-technology.com>

- 30 mg/g OR UO = 30-300 mg/g OR UO > 300 mg/g)
- o Stage 5 (ESRD): GFR < 15 AND (UO < 30 mg/g OR UO = 30-300 mg/g OR UO > 300 mg/g)
- **REQ3:** The interpreter of kidney failure states per a patient must be displayed at only one stage.

**Use Case based Specification Model.** The use case diagram of KFD subsystem is shown in Fig. 2. It consists of three modules as demonstrated as use cases UC001-UC003 that covers all KFD functionality including GFR calculation, interpretation and display that correspond to requirements REQ1-REQ3, respectively. The detailed functionality of each use case is represented in the form of use case description as shown in Fig. 3. This figure demonstrates a brief description of use case UC001 that describes the behaviour of function GFR calculation. It explains the sequence of the step for use case UC001 including *main flow of events*, *alternative of events* and *exception flow* of each use case. The flow of events demonstrates the main step-by-step of GFR calculation. Furthermore, the input for GFR calculation including *Gender*, *Age* and *SCr* used to calculate *GFR* is also indicated in the corresponding step. After the GFR calculation, the GFR stage is interpreted as an output of this functionality. Based on the input and output, the relevant variables affected the change of test cases can be analysed and stored in the XML dictionary document for generating test data.

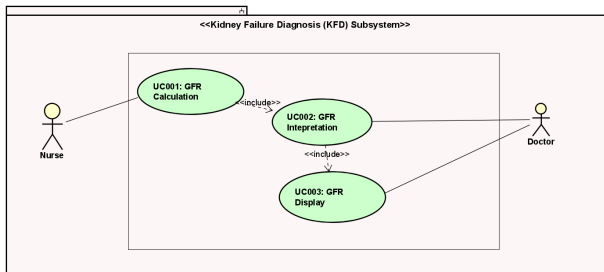


Fig. 2. Use case model for KFD subsystem

Use Case#	UC001: GFR Calculation
Pre-conditions:	System shows the menu to record GFR
Primary Actor:	Nurse
Flow of Events:	1 The system shows the GUI to input data. 2 The nurse inputs medical record ( <i>Gender</i> , <i>Age</i> and <i>SCr</i> ). 3 The nurse saves the medical record. 4 The system calculates GFR [E1] [E2] [E3]. 5 The system interprets GFR stage via the use case UC002, <i>GFR Interpretation</i> . 6 The system displays GFR stage via the use case UC003, <i>GFR Display</i> .
Alternative of Events:	...
Exception Flows:	1 [E1] If the nurse inputs the variables ( <i>Gender</i> , <i>Age</i> and <i>SCr</i> ) that are not in the correct range, the system will display "Error GFR calculation". 2 [E2] If the system maps wrong input of variables, the system will display "Wrong Interpretation". 3 [E3] If the system maps input that is not in the correct range, the system will display "GFR stage does not match".
Version	01

Fig. 3. Example of use case description of UC001 (Version 1)

Considering each use case diagram of the KFD subsystem, variables from the data dictionary of the system are introduced and used as input data, as shown in Table II. This table contains *variable names*, *variable types* and *data range* values. The type of variable and the range value of data for each variable are determined for the purpose of creating the test data such as

valid variables of *Gender* are *F* and *M*, variable *Age* contains the correct data range values: *0-120*, which is in the range of data interest.

TABLE II. DATA DICTIONARY OF UC001 (VERSION 1)

Variable Name	Description	Data Type	Data Range
Gender	Gender of patients	Varchar(1)	F, M
Age	Age of patients	Integer	0-120
SCr	Serum Creatinine	Float	0.0-10.0
UO	Urine Albumin	Integer	0-∞
GFR	Glomerular Filtration Rate	Integer	0-∞
Stage	Kidney Failure Stage	Varchar(255)	Stage 1-Stage 5

**Requirement Change Requests.** There were the changes of the detailed requirement in the first version of use case based specification model from the user that were recorded in the requirement change request form as shown in Table III. The user requested to change the formula for GFR calculation as identified in **CH01**. This change request affected the functionality of requirement **REQ1**. In the first version, only a single formula was used to calculate GFR for all ages. In the changed version, five formulas are used for different ages instead as demonstrated in Table IV. The GFR calculation is measured separately for an adult patient whose age is greater than and equal to 18 with different *Sex* and *SCr* and a child patient who has age less than 18. Moreover, the factor *Height* is also included in the changed formula for the child patient. There was also change request **CH02** that results in the change in requirement **REQ2** in which variable *Gender* was renamed to be *Sex*. In the last change request **CH03**, the data type of GFR were also changed from integer to float. The relevant variables in the changed version are analysed as shown in Table V.

TABLE III. THE CHANGE REQUEST OF KFD SUBSYSTEM

Change ID	Change Description	Affected Requirement ID	Affected Use Case ID
CH01	Change the formula for GFR Calculation	REQ1	UC001
CH02	The name of variable change from gender to sex based on the table of GFR formula	REQ2	UC003
CH03	Change data type of <i>GFR</i> to decimal	REQ3	UC002

TABLE IV. THE CHANGED GFR FORMULA

Age	Sex	SCr	Formula
≥ 18	F	≤ 0.7	$GFR=144 \times (SCr/0.7)^{-0.329} \times (0.993)^{Age}$
		> 0.7	$GFR=144 \times (SCr/0.7)^{-1.209} \times (0.993)^{Age}$
	M	≤ 0.9	$GFR=141 \times (SCr/0.9)^{-0.411} \times (0.993)^{Age}$
		> 0.9	$GFR=141 \times (SCr/0.9)^{-1.209} \times (0.993)^{Age}$
< 18	F/M		$GFR=0.413 \times Height(cm)/SCr(mg/dL)$

Note: F=Female, M=Male

Fig. 4 and Table V describe the use case description and its corresponding details of data dictionary in the changed version respectively. Based on the requirement change, there is a change in variable *Age* in which the data range value of this variable is changed from *0-120* to two range values that are *0-17 (<18)* and *18-120 (≥ 18)*. As the value of variable *SCr* is calculated based on the value of variables *Sex* and *Age*

as shown as the different equations in Table IV, This affects the value of this variable to be changed. The range value of factor *SCr* is divided into two main criteria for female and male which is ( $\leq 0.7$ ,  $> 0.7$ ) and ( $\leq 0.9$ ,  $> 0.9$ ) respectively. Furthermore, as the height of patients is included into the changed formula, variable *Height* with its range value (0-300) is added into the new version of data dictionary. Finally, the type of variables *GFR* is changed from integer to floating point as to increase more accurate and precise result interpretation.

Use Case#	UC001: GFR Calculation
Pre-conditions:	System shows the menu to record GFR
Primary Actor:	Nurse
Flow of Events:	<ol style="list-style-type: none"> <li>1 The system shows the GUI to input data.</li> <li>2 The nurse inputs medical record (<i>Gender, Age, Height and SCr</i>).</li> <li>3 The nurse saves the medical record.</li> <li>4 The system calculates GFR by using the formula depends on <i>Gender</i> and <i>Age</i> [E1] [E2] [E3].</li> <li>5 The system interprets GFR stage via the use case <i>UC002, GFR Interpretation</i>.</li> <li>6 The system displays GFR stage via the use case <i>UC003, GFR Display</i>.</li> </ol>
Alternative of Events:	...
Exception Flows:	<ol style="list-style-type: none"> <li>1 [E1] If the nurse inputs the variables (<i>Gender, Age, Height and SCr</i>) that are not in the correct range, the system will display "Error GFR calculation".</li> <li>2 [E2] If the system maps wrong input of variables, the system will display "Wrong Interpretation".</li> <li>3 [E3] If the system maps input that is not in the correct range, the system will display "GFR stage does not match".</li> </ol>
Version	02

Fig. 4. Use case description of UC001 (Version 2)

TABLE V. DATA DICTIONARY OF UC001 (VERSION 2)

Variable Name	Description	Data Type	Data Range
<i>Sex</i>	<i>Gender of patients</i>	<i>Varchar(1)</i>	<i>F, M</i>
<i>Age</i>	<i>Age of patients</i>	<i>Integer</i>	$< 18, \geq 18$
<i>SCr</i>	<i>Serum Creatinine</i>	<i>Float</i>	$\leq 0.7, > 0.7, \leq 0.9, > 0.9$
<i>Height</i>	<i>Height of patients</i>	<i>Integer</i>	<i>0-300</i>
<i>UO</i>	<i>Urine Albumin</i>	<i>Integer</i>	<i>0-∞</i>
<i>GFR</i>	<i>Glomerular Filtration Rate</i>	<i>Float</i>	<i>0.0-∞</i>
<i>Stage</i>	<i>Kidney Failure Stage</i>	<i>Varchar(255)</i>	<i>Stage 1-Stage 5</i>

2) *Online nursing assessment (ONA) subsystem*: ONA subsystem is used to record the clinical information of the patient in order to be used to evaluate the symptoms of the patient and consider the treatment for the doctor including medicine dispensing to patients and accurate treatment.

**Requirement.** The following are the requirements of this subsystem:

- **REQ1:** Nurses shall record nursing assessment data such as patient severity (*Triage*), service type (*Visit-Type*), service status (*VisitStatus*), *allergy, pregnancy, and lactation* status.
- **REQ2:** Nurses shall record the clinical information of the patient including vital signs (V/S), which includes pulse rate (*pulse*), respiratory rate (*RR*), body temperature (*Temp*) and blood pressure (*BP*), as well as weight (*weight*), height (*height*) and waistline (*waistline*) for patients. The normal range value of each clinical information is shown as follows.
  - *Pulse:* 60-100 time/min.
  - *PR:* 12-18 time/min.
  - *Temp:* 36.5-37.5 °C
  - *BP:* 90/60-120/80 mmHg

- **REQ3:** The system shall automatically calculate Body Mass Index (*BMI*) for patients. The formula is calculated as *BMI = Weight divided by square meter of height*.
- **REQ4:** The level of nutrition is divided into five levels as follows [15]:
  - *BMI < 18.5:* Underweight
  - *BMI ≤ 18.5 and BMI ≥ 24.9:* Normal weight
  - *BMI ≤ 25.0 and BMI ≥ 29.9:* Pre-obesity
  - *BMI ≤ 30.0 and BMI ≥ 34.9:* Obesity class I
  - *BMI ≤ 35.0 and BMI ≥ 39.9:* Obesity class II
  - *BMI ≥ 40.0:* Obesity class III

**Use Case based Specification Model.** The use case diagram corresponding to the requirements described in the previous section is shown in Fig. 5. There are two actors specified in this use case, the nurse and doctor. Two main use cases are identified, *UC01: Assessment Record* and *UC02: Vital Sign Record*. Use case *UC01* implements requirement **REQ1** regarding nursing assessment record, whereas requirement **REQ2** regarding operation recording the clinical information of the patient is implemented by use case *UC02*. After recording such clinical information, the calculation and interpretation operation is performed based on requirements **REQ3** and **REQ4**. Therefore, three use cases that are *UC021: BMI Calculation*, *UC022: Nutrition Interpretation* and *UC023: Nutrition Display* are included by use case *UC02*.

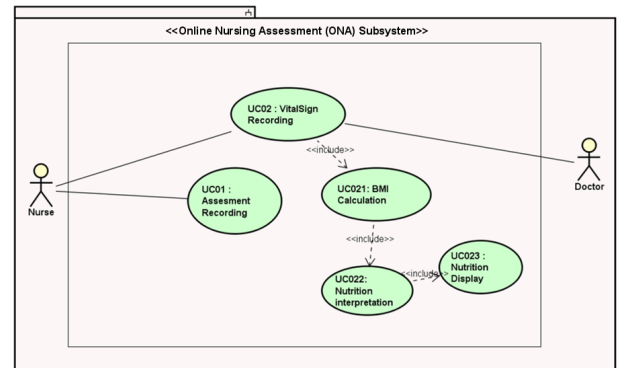


Fig. 5. Use case model for ONA subsystem

Fig. 6 shows the use case description of use case *UC01* (Version 1). This figure specifies the details of use case including the sequence of each step of use case *UC01* including *Main Flow of Events*, *Alternative of Events* and *Exception Flows*. As can be seen in this figure, the description of use case *UC01* is explained regarding the functionality of nursing assessment record. Considering the flow of events in this table, variables *Triage*, *VisitType*, *VisitStatus*, *Allergy*, *Pregnancy*, *Lactation* and *PainScore* are indicated as an input data. These variables and their data structure are defined in the data dictionary as the form of XML file format. This file is used later to generate test data. These variables including their data structure and possible data range are shown in Table VI. There are 19 variables included in the first version of data dictionary used by the functionality of ONA subsystem. For example, variable *VisitType* that has data type as a string for recording the reason of visit. This variable contains nominal data value to be *Self*, *Relative* and *Other*. Another example is variable *Triage* that



determines the priority of patients’ treatments based on the severity of their condition. This variable has ordinal data range value of a character from 1 to 5.

Use Case#	UC01: Assessment Recording
Pre-conditions:	Select the patient from the system
Primary Actor:	Nurse
Flow of Events:	1 The system displays the assessment recording window. 2 The nurse inputs assessment information including severity ( <i>Triage</i> ), type of service ( <i>VisitType</i> ), service situation ( <i>VisitStatus</i> ), <i>Allergy</i> , <i>Pregnancy</i> , <i>Lactation</i> and <i>PainScore</i> of patients. 3 The nurse clicks “Save” button to save assessment information [A1].
Alternative of Events:	1 [A1] The nurse clicks “Cancel” button in the case of cancellation [E1].
Exception Flows:	1 [E1] The system will clear screen.
Version	01

Fig. 6. Use case description of UC01 (Version 1)

TABLE VI. DATA DICTIONARY OF UC01 (VERSION 1)

Variable Name	Description	Data Type	Data Range
Sex	Gender of patients	Varchar(1)	F, M
Age	Age of patients	Integer	1-120
Pulse	Pulse Rate	Integer	0-201
RR	Respiratory Rate	Integer	0-120
Temp	Body Temperature	Integer	35-46
BP_upper	Systolic blood pressure	Integer	0-350
BP_lower	Diastolic blood pressure	Integer	0-350
Weight	Weight of patients	Integer	0-200
Height	Height of patients	Integer	0-300
WL	Waist of patients	Integer	5-80"
BMI	Body Mass Index	Float	18.50-30.00
VisitType	Service Type	Varchar(255)	Self, Relative, Other
VisitStatus	Service Status	Varchar(255)	Walk, Wheelchair, Other
Triage	Triage Level	Varchar(1)	1-5
Allergy	Allergy Status	Varchar(1)	Y, N
Preg	Pregnancy Status	Varchar(1)	Y, N
Lactation	Lactation Status	Varchar(1)	Y, N
Pain Score	Pain Level	Integer	1-10
NT Level	Nutrition Level	Varchar(2)	L1-L5

**Requirement change requests.** After we analysed the requirement change request, we discovered that this change of requirements affected the functionality of requirement as shown in Table VII. There are two types of changes, deleting and adding some requirements respectively. This impacted the use case description and data dictionary to be changed as shown in Fig. 7 and Table VIII, respectively. According to change request **CH01**, *Pregnancy* and *Lactation* of patient are excluded from inputs of assessment information in the second version of the use case description. This affected the variables *Preg* and *Lactation* specified in the first version of data dictionary were deleted. Change request **CH02** introduces the new functionality into the subsystem to display the nursing record information after assessment recording functionality, the new use case (*UC011*) is added and extended from *UC01*. This added use case results in variable *DateTime* to be added in order to record the patient information based on the selected date. This variable contains nominal data range values *CD* and *UD* for the current date and updated date respectively. Furthermore, some variables changed their data range value. For example, from change request **CH03**, variable *Pulse* were separated from one partition (0-201) into two partitions as 60-100 and 90-130 for different age groups respectively. Change request **CH04** affected the name of variable *NT Level* to be renamed to *BMI Level* according to the international standard name based on WHO.

TABLE VII. THE CHANGE REQUEST OF ONA SYSTEM

Change ID	Change Description	Affected Requirement ID	Affected Use Case ID
CH01	Delete pregnancy and lactation details	REQ1	UC01
CH02	The system can display the nursing record information based on the selected date (current date: <i>CD</i> and updated date: <i>UD</i> )	REQ2	UC011
CH03	Adjust the patient’s normal range including <i>Pause</i> : 90-130 times per minute for children aged 18 years old and 60-100 times per minute for adult aged over 18 years old	REQ3	UC02
CH04	Rename the nutrition level display to BMI level	REQ4	UC021
	Change data type of <i>Weight</i> and <i>Height</i> to decimal	REQ4	UC023

Use Case#	UC01: Assessment Recording
Pre-conditions:	Select the patient from the system
Primary Actor:	Nurse
Flow of Events:	1 The system displays the assessment recording window. 2 The nurse inputs assessment information including severity ( <i>Triage</i> ), type of service ( <i>VisitType</i> ), service situation ( <i>VisitStatus</i> ), <i>Allergy</i> and <i>PainScore</i> of patients. 3 The nurse clicks “Save” button to save assessment information [A1].
Alternative of Events:	1 [A1] The nurse clicks “Cancel” button in the case of cancellation [E1].
Exception Flows:	1 [E1] The system will clear screen.
Version	02

Fig. 7. Use case description of UC01 (Version 2)

### C. CIA Algorithm

To achieve a better understanding of our impact analysis framework as described in Section III-A, the detailed algorithm is explained in this section. This is demonstrated with using our experienced case study, KFD subsystem as an example.

**Step 1: Variable Extraction.** After the change request is identified by the change analysis, variables specified in the use case diagram are extracted. As variables and their data structure are normally defined in the data dictionary, the XML of data dictionary that is extended from the XML of conformed use case is used as a source for this variable extraction in our developed prototype tool. The variable information in the XLM file including *use case id*, *variable name*, *variable type*, *variable range value*, *variable number* and *variable order* are extracted before it is analysed for the level of change impact.

Fig. 8 demonstrates an example of the changed version of the XML of data dictionary that conforms to the changed version of use case description. The information of variables *Sex*, *Age*, *Height* and *SCr* encoded in this XML file is extracted. Especially, the data structure that are *variable type* and *variable range value* from the extracted information is normally used as information to generate test cases or test data. For instance, considering variable *Age* which has data type as an integer, there are four possible test cases to be considered, the age in the range of 0-17, 18-120 and out of the range that are less than 0 and greater than 120.

**Step 2: Two Versions of Use Case File Comparison.** This step performs consistency checking in order to identify the variables affected from the change of requirement. Two versions of the extracted variable information between before-change and after-change of the used case diagram from the previous step are compared and analysed. This results in the

TABLE VIII. DATA DICTIONARY OF UC01 (VERSION 2)

Variable Name	Description	Data Type	Data Range
Sex	Gender of patients	Varchar(1)	F, M
Age	Age of patients	Integer	1-120
Pulse	Pulse Rate	Integer	60-100 for children aged 18 years old, 90-130 for adult aged above 18 years old
RR	Respiratory Rate	Integer	0-120
Temp	Body Temperature	Integer	35-46
BP_upper	Systolic blood pressure	Integer	0-350
BP_lower	Diastolic blood pressure	Integer	0-350
Weight	Weight of patients	Float	0.0-200.0
Height	Height of patients	Float	0.0-300.0
WL	Waist of patients	Integer	5-80''
BMI	Body Mass Index	Float	18.50-30.00
VisitType	Service Type	Varchar(255)	Self, Relative, Other
VisitStatus	Service Status	Varchar(255)	Walk, Wheelchair, Other
Normal	Normal Status	Boolean	1=true,0=false
Datetime	Recorded Date Time	Varchar(2)	CD, UD
Triage	Triage Level	Varchar(1)	1-5
Pain Score	Pain Level	Integer	1-10
BMI Level	BMI Level	Varchar(2)	L1-L5

```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
3  <xs:project>
4  <xs:usecase id="UC001" name="GFR Calculation"/>
5  <xs:ucelement name="input">
6  <xs:complexType>
7  <xs:sequence>
16  <xs:element id='2' name="Age">
17  <xs:simpleType>
18  <xs:restriction base="xs:integer">
19  <xs:minLength id='1' value="0"/>
20  <xs:maxLength id='1' value="17"/>
21  <xs:minLength id='2' value="18"/>
22  <xs:maxLength id='2' value="120"/>
23  </xs:restriction>
24  </xs:simpleType>
25  </xs:element>
42  </xs:sequence>
43  </xs:complexType>
44  </xs:ucelement>
45 </xs:project>
46 </xs:schema>

```

Fig. 8. The XML file of UC001 (Version 2)

cause of changes to be identified and analysed. There are five patterns of atomic changes for determining the cause of changes that are 1) change of variable name, 2) change of variable type, 3) change of variable range value, 4) change of number of variables and 5) change of variable order. The possible cause of change patterns affects to the addition, deletion and modification of target variables.

Considering the changed GFR calculation formulas in KFD subsystem as shown in Table IV, for example, the comparison of two versions of use case enables us to discover the causes

of variable changes. 1) the modification of variable value is detected as the range value of variable *Age* was changed from 0-120 to 0-17 and 18-120. Furthermore, 2) the addition of variable is identified as variable *Height* for GFR calculation. The before-changed test cases directly affected from these variable changes are later discovered to be re-generated.

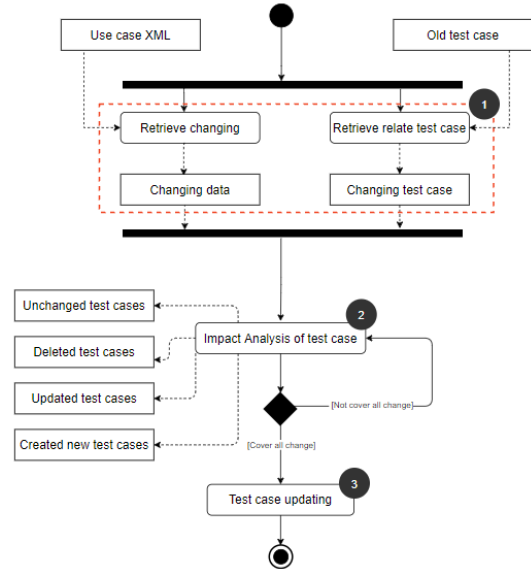


Fig. 9. Procedure of impact analysis of test cases

**Step 3: Impact Analysis of Test Cases.** After the variable affected by the change is identified and considered, the change impact of test cases is analysed. Fig. 9 shows the procedure of the change impact of test cases that is divided into three steps.

(1) *Retrieve Inputs for Impact Analysis:* this step retrieves and prepares the necessary inputs used for analysing the change impact of test cases. Two sets of data are prepared in this step: the information of changed variables and a set of before-changed test cases. To prepare this, the information of the changed variables together with the cause of changes is input from the previous step, whereas a set of before-changed test cases is retrieved from the test case database. This results in the information of changing data in affected variables and changing test cases to be generated and used in the next step.

(2) *Impact Analysis of Test Cases:* the information of changed variables prepared from the previous steps is used to analyse and validate the impact on the before-changed test cases. This results in these test cases being classified into four groups that are *no-change*, *delete*, *update* and *create* categories. Our developed tool also supports this step to analyse effectively the change impact from the multiple various change patterns occurring in parallel at the same time. The patterns for identifying the action of effect on test cases based on type of change impact are summarised in Table IX. The percentage of estimated reusability is also calculated for each pattern.

Considering in the case of 1) the change of variable name, the impact of this case affects the corresponding test cases of this variable to be not changed. Thus, they can be totally reused (100%). For 2) the change of data type, the existing

TABLE IX. SUMMARY OF FIVE PATTERNS OF AUTOMATIC CHANGES WITH EFFECT ON TEST CASES

Type of Impact		Action				% Est. Reuse
		No Change	Update	Delete	New	
Change of variable name	-	×				100
Change of data type	-		×			100
Change of number of variables	Add			×	×	0
	Delete			×	×	0
Change of variable value	Add	×	×		×	50
	Delete	×	×	×		50
	Update		×			100
Change of variable order	-			×	×	0

test cases can be 100% reused with only updating their data type. 3) the change of number of variables, in both cases of adding and deleting the variables, results in all existing test cases not to be totally reused. This is because of the Cartesian product used for generating test cases needs to recalculate the multiplication of all variable range value to support the deletion or addition of variable. This affects the existing test cases corresponding to these variables to be the deleted and the new test cases to be generated. 4) The change of range value of variables is divided into three possible cases (4.1) adding a new range value in which the valid partition is added or extended from the other valid partition. This results in all test cases generated from this partition are totally reused as there is no change in the valid partition. However, this case causes the intersection of range values between the addition valid partition and the other invalid partition. This, therefore, results in the range value of affected invalid partition to be reduced by removing the overlapped value. As a result of this, the existing test cases affected from this reduced invalid partition to be updated and new test cases are generated from the additional valid partition. Similar to this, (4.2) deleting existing range value affects the existing test cases generated from unchanged partition to be totally reused and those generated from deleted partition to be deleted. Furthermore, this case affects the range value of invalid partition to be expanded to cover the deleted partition. Therefore, this results in the test case corresponding to this to be updated. In cases (4.1) and (4.2), the percentage of reusable test cases is estimated to be 50%. In the case of (4.3) updating existing range value that one partition is split into two partitions, this causes the existing test cases to be updated (100% of reusability). Finally, we have discovered that the change number of variables as adding new variables and deleting existing variable affects the order of variable. Thus, 5) the change of variable order affects some existing test cases to be deleted and new test cases to be generated.

(3) *Test Case Modification*: the before-changed test cases are considered to be modified based on the result of impact analysis that is classified into the impact group as mentioned earlier. For the test case that is classified into *no-change* group, it can be totally reused for the new version. However, the before-changed test cases have their value updated or are deleted when the result of impact analysis is classified into the *update* or *delete* groups respectively. If the impact analysis of test cases results in the *create* group, a new test case is generated. All modified and created test cases are generated by using CCTM algorithm.

Table X shows an example of the variable partitions for test case generation after analysing the impact of test cases. As can be seen in this table, based on the requirement change request of GFR Calculation module mentioned before, the range of variable *Age* is changed from one partition (partition 5 in variable before change) for the valid test case, 0-120 years old, to two partitions (partitions 5 and 6 in variable after change) for the valid test case, 0-17 and 18-120 years old. This causes the updated test case to be improved by creating a new version of test cases corresponding to the updated partition as shown in Table XI. The new test cases corresponding to the separated partitions after the change (0-17 and 18-120 years old) are re-generated as in test cases 2-3 (as shown in the highlighted row in the table). Furthermore, from the no-change group of impact analysis, there are two original (unchanged) test cases that are totally reused in the new set of test cases, test cases 1, 4-5 respectively.

TABLE X. EXAMPLE OF UPDATING THE VALUE OF A VARIABLE AFTER CHANGING

Variables before change					Variables after change				
Class	Var.	Min.	Max.	Type	Class	Var.	Min.	Max.	Type
1	Sex	F	F	Valid	1	Sex	F	F	Valid
2		M	M	Valid	2		M	M	Valid
3		N/A	N/A	Invalid	3		N/A	N/A	Invalid
4	Age	-∞	-1	Invalid	4	Age	-∞	-1	Invalid
5		0	120	Valid	5		0	17	Valid
6		121	∞	Invalid	6		18	120	Valid
					7		121	∞	Invalid

TABLE XI. EXAMPLE OF TEST CASES AFTER CHANGING

#TC	Class	Variable name		Sequence		Type
		Sex	Age	Sex	Age	
1	1,4	F	-4	1	2	Invalid
2	1,5	F	12	1	2	Valid
3	1,6	F	55	1	2	Valid
4	1,7	F	176	1	2	Invalid
...	...	...	...	...	...	...

**Step 4: Test Case Generation.** The approach performs test case generation when test cases need to be recreated. CCTM, the combination technique of ECP and CTM, is applied to cover generated test cases with all functions as required. The main steps are as follows:

**Step 4.1: Analyse Requirements using a Classification Tree:**

The classification tree technique basically enables a set of requirements in the specification to be classified as a branch in the tree. The top of the tree as a root is the main use case in the system before the sub use case is considered to extend the root to create the sub-tree. A lower level of the tree, variables of the use case are identified as terminal classification. The terminal class is the range value of the variable and identified at the lowest level as a leaf of the classification tree. There are seven principle combination patterns proposed in [16]. One example of these combination patterns that was revealed by the tool is shown in Fig. 10. The tool detected two redundant classification trees that were built to support the calculation of two age groups as shown in Fig. 10a and 10b respectively. These redundant terminal classes were eliminated by CTM enabling in our tool. The terminal class values were combined into the same tree as shown in Fig. 10c. This influences the

benefit to test case generation step in which the potential redundancy of generated test cases is reduced.

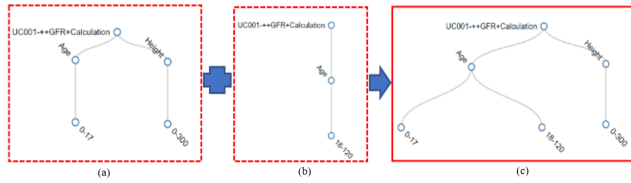


Fig. 10. Example of a combined use case

Fig. 11 shows an example of classification tree of GFR calculation module after the analysis of changes. It contains the variables and their corresponding possible range value visualising in terminal classification (parent node) and terminal class (leaf node) respectively. The range value of all relevant variables demonstrated in this tree can be explained as follows: *Sex* = *F*, *M*, *Age* = *0-17*, *18-120*, *Height* = *0-300* and *SCr* = *0-0.9*, *1.0-10.0*. These are considered to create equivalence class partitioning with ECP technique for the next step.

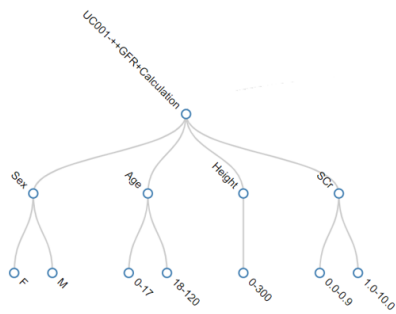


Fig. 11. The classification tree of GFR module

**Step 4.2: Generate Test Case from the Equivalence Class Partitioning:** This step generates the test case from the terminal class by using the ECP technique enabled by CCTM method. The value of variable indicated in this terminal class is considered to create the partition. In our developed prototype, the partition of changed variables is generated automatically resulting as in Fig. 12. This figure illustrates the partition created from the terminal class in the updated classification tree demonstrated in Fig. 11. Variable *Age* is divided into four partitions (two valid partitions and two invalid partitions) corresponding to the changed version of variables. To generate test cases, ECP performs the Cartesian product of multiplying a data value selected from all partitions.

#### IV. PROOF OF CONCEPT

##### A. Tool Development

A prototype tool was developed to demonstrate the effectiveness of the CIA framework for test cases explained in Section III-A. It is a Java based web application implemented in the Node.js 8.9.4<sup>2</sup>, JavaScript run-time environment which is widely used as a standard for large-scale application. MySQL

<sup>2</sup><https://nodejs.org/en/about/>



Fig. 12. Example of equivalence class partitioning

version 5.7.22<sup>3</sup> is used to keep the data of use case version, variables and test data for test suite export.

Fig. 13 demonstrates an example of our developed prototype tool. Fig. 13a is the screen showing two versions (before-change and after-change) of XML file of use case diagram after they are uploaded before the consistency checking performs. Then, the tool performs variable analysis of each version of the XML file. All variables defined in the use case are analysed. This includes variable name, variable type and variable range value as demonstrated in Fig. 13b. Fig. 13c shows the screen of impact analysis result. It shows the summary of impact analysis which is divided into two parts. The first part reports the number of changes classified and grouped into five patterns of atomic changes as we proposed in the framework as mentioned in Section III-A. The reuse percentage of test cases affected from the change is also calculated and reported on this screen. The latter part reports the number of test cases affected from the changes grouped by the causes of changes that are *No Change*, *Delete*, *Update*, *New (Create)*. Finally, test case generation is performed when test cases need to be recreated. CCTM technique is used for test case generation. This results in the classification tree and its corresponding equivalence class partitions is created and analysed as shown in Fig. 11 and 12, respectively before generating the test cases and test data.

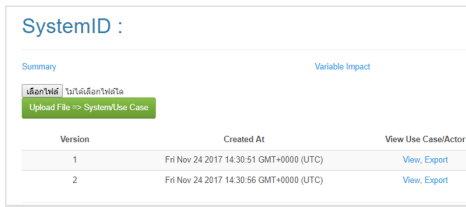
##### B. Tool Validation

Before evaluating the the effectiveness and efficiency of our proposed approach, we validated the correctness of the tool to confirm whether all functionalities of the tool demonstrated in Section IV-A perform correctly according to the framework presented in Section III-A. Therefore, we conducted experimental validation aiming to answer the following question.

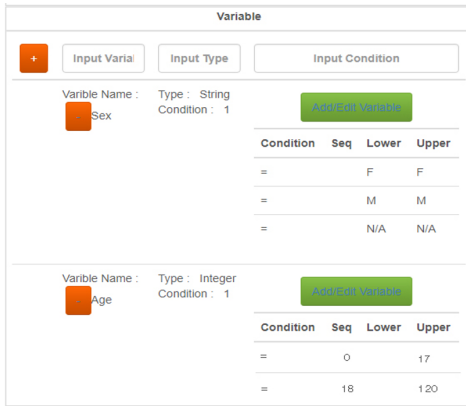
**EQ1:** *Are all functionalities implemented in the tool implemented and do they perform correctly according to the proposed approach?*

To accomplish this, three test scenarios corresponding to four steps of the CIA algorithm described in were designed as shown in Table XII including 1) *TS-01 : Validate Comparing Two Versions of Variable Function*, 2) *TS-02: Validate Test Case Analysis Function* and 3) *TS-03: Validate CCTM function*. Test scenario *TS-01* aims to validate the comparison

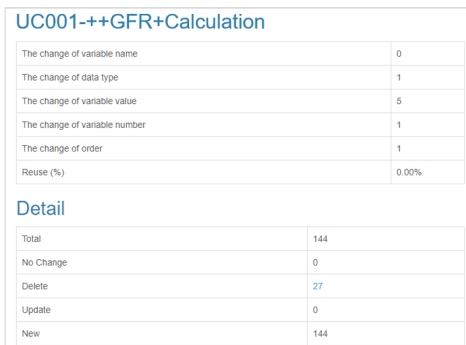
<sup>3</sup><https://www.mysql.com/>



(a) screen of XML version



(b) variable management



(c) impact analysis

Fig. 13. Example of the tool screens

operation of versions of variable XML function according to Steps 1 and 2 in CIA algorithm, after two versions of the XML data dictionary containing the data structure of variables are uploaded. The change of variables is compared based on the proposed five patterns. For test scenario *TS-02* aiming at the validation from Step 3 of the algorithm, impact analysis of test case function is validated after the comparison of change of variable based on five patterns. Test scenario *TS-03* is used to validate the operation of classification tree generation by CCTM technique. As the main objective of CCTM technique is to reduce the duplicated test cases by merging the duplicated classification tree caused from the redundant requirements, we also created the scenario for validating this tree merging operation.

Table XIII shows the result of testing. Test cases in test scenario *TS-03* enabled us to reveal the fault in our developed tool. As can be seen in this table, seven test cases were designed from merging two classification approach proposed in [16] and used to validate our merging classification tree

TABLE XII. THE RESULT OF TOOL TESTING

Test Scenario	Corresponding Steps	Result	Revision
TS-01 Validate Comparing Two Versions of Variable Function	Steps 1 and 2	Fail	Pass
TS-02 Validate Test Case Analysis Function	Step 3	Pass	-
TS-03 Validate CCTM Function	Step 4	Fail	Pass

function. We discovered the fault in test case *TC-037* regarding the case that two use cases have the same variable name but some range value of variable in one use case is overlapped with that of the other. The expected result of this test case should result in three sub-trees to be created after merging tree as the intersection of variable range between two use cases and two exteriors of variable range in both use cases. The test result revealed that our tool created only two sub-trees instead of three sub-trees after merging tree operation. This led us to fix the code to perform the correct operation of this merging tree case. Furthermore, the fault was also found in test scenario *TS-01* caused by the problem of XML data dictionary that could not trace the change of variable. Thus, we redesigned the XML data dictionary and retested this function. All revision tests were passed as we expected.

TABLE XIII. THE TESTING RESULT OF TEST CASE IN TEST SCENARIO *TS-03*

Test Cases	Objective	Pass/Fail
TC-031	The integration of two use cases in which the other use case has no information	Pass
TC-032	The integration of two use cases that have the same terminal sub-tree	Pass
TC-033	The integration of two use cases in which these two use cases have the same variable name and value range	Pass
TC-034	The integration of two sub-use cases in which share the same use case (parent) but have the different variable name and value range	Pass
TC-035	The integration of two use cases in which these two use cases have the same variable name but different value range	Pass
TC-036	The integration of two use cases in which these two use cases have the same variable name but all value range of variable in one use case is in that of the other	Pass
TC-037	The integration of two use cases in which these two use cases have the same variable name but some value range of variable in one use case is overlap with that of the other	Fail

Based on experimental validation, the testing result of test scenarios and cases confirms the accuracy and coverage of functionality in which all functionalities corresponding to the steps in CIA algorithm were implemented in the tool and they performed correctly according to the proposed approach.

### C. Tool Evaluation

The effectiveness and efficiency of our proposed approach were evaluated by using two real-world case studies, 1) KFD and 2) ONA subsystems, explained in Section III-B. The result of change impact analysis and test case regeneration from these two case studies generated automatically from the tool was compared with that of manual operation. Furthermore, we evaluated the satisfactory level of the developed tool from the target user for potential application in the future. The aim of tool evaluation is to answer the following evaluation questions.

**EQ2:** How error-prone can be identified from the comparison between manual CIA and automated tool?

**EQ3:** How the impact analysis result can influence the level of reusability of existing use cases?

**EQ4:** How do potential users assess a satisfactory result of CIA from the automated tool compared with the manual CIA?

1) *Effectiveness evaluation:* To answer **EQ2**, the comparison between the actual result of impact analysis created by the automatic tool and the expected result calculated by experts was conducted. The precision, recall and F-measure computation were calculated by comparing the result produced by the manual CIA and automated tool. The computation metrics were adapted from [17] as follows.

$$Precision = \frac{|\{Expert\ Identified\} \cap \{Tool\ Identified\}|}{|\{Tool\ Identified\}|} \times 100 \quad (1)$$

$$Recall = \frac{|\{Expert\ Identified\} \cap \{Tool\ Identified\}|}{|\{Expert\ Identified\}|} \times 100 \quad (2)$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

Table XIV demonstrates the result of impact analysis. We compared the number of actual test cases after performing impact analysis generated automatically by the tool with the number of expected test cases calculated manually by experts. Considering the calculated F-measure with precision and recall of KFD case study, the accuracy of the automatic tool performing with this case study is very high. This is because the KFD case study is not a complex case study containing basic changes such as renaming a variable name and change the variable type which did not affect much to the test cases. However, as ONA is a more complex case study containing change of variable range value that was not specified in KFD case study, the accuracy of the automatic tool performing with ONA case study in that pattern is low. This also affected the accuracy of analysing pattern change of variable order to be low. After we revealed the tool, we discovered that that tool excluded the additional range value affected from the adjusted value of variables. This led us to correct the tool to include this case when the tool performs test case generation. After fixing the tool, the number of regenerated test cases by the tool is the same as that of expected test cases hand-operated by experts. The result produced by the corrected tool is shown in Table XV.

2) *The level of reusability:* To answer question **EQ3** regarding the level of reusability, we considered the impact analysis result of test case evaluated from two case studies as shown in Table XV. This table shows the total number of test cases generated after change impact analysis of all use cases of two case studies obtained from the developed tool. The amount of reuse equation adapted from [18] is used to assess the level of reusability of test cases as shown in Eq. (4).

$$\%Reuse = \frac{\# of\ Reused\ Test\ Cases}{\# of\ Total\ Test\ Cases} \times 100 \quad (4)$$

In KFD subsystem, 185 test cases were generated. Based on this number, three types of changes are discovered by the tool. (1) Changing variable name in use case *UC003* resulted in 6

test cases of the first version to be totally reused. Furthermore, 35 test cases to be updated due to changing data type in use case *UC002*. Considering the case of (3) adding a new variable (as *Height*) in use case *UC001*, 27 test cases of version 1 were deleted and 144 of new test cases were generated for version 2. Furthermore, the reusability level of test cases was also measured from the no-change and update group of test cases resulted from the impact analysis. Around 22% of existing test cases for KFD subsystem testing can be reused in the new version of test cases. The reusability rate is low. This is because of the effect from the addition of new variable. As the Cartesian product of multiplying all variable range value is used to generate test cases, all existing test cases affecting to this change were not totally reused. All existing test cases regarding to these variable to be deleted and new test cases to be created by recalculating the Cartesian product instead.

In ONA subsystem, five types that cover all possible changes were identified by the tool. (1) Changing variable name in use case *UC023* affected to 21 test cases of the first version to be totally reused. (2) Changing data type in use case *UC023* resulted in 9 test cases to be updated. Furthermore, (3) adding a new variable as in use case *UC011* caused 6 of new test cases to be generated for after-changed version. (4) Deleting an existing variable in use case *UC01* resulted in 2,592 of test cases to be deleted and 288 of new test cases to be created for version 2. Lastly, there was (5) the addition of variable range value occurred in use case *UC02* of this case study, in which the first case study did not have, was revealed by the tool. This resulted in 4,374 test cases in version 1 to be reused 100 percent (no change), another 4,374 test cases in this before-changed version to be updated and 17,496 of new test cases to be generated. The reusability level of test cases for ONA subsystem is approximately 33% of existing test cases being reused in the new version. This is because there was change of variable range value. This affects half of existing test cases related to this change effect can be totally reused.

3) *Satisfaction evaluation:* To answer **EQ4**, we evaluated the satisfaction of our proposed approach and tool from the target user. Practical specialists who have experience in HIS system development for more than five years were selected as the target user. This was also chosen from the wide range of the role in the development team that are 1 development manager, 2 software testers and 1 developer. The evaluation process started with the tool being demonstrated and trained to specialists before they used the tool with the prepared case study. The satisfactory level with the tool was evaluated by using the evaluation form. Likert scale ranking from strongly agree (5) to strongly disagree (1) indicated a satisfactory level in each question of the evaluation form.

Table XVI demonstrates the question used in a satisfactory evaluation form and the evaluation analysis result is shown in Fig. 14. Most specialists strongly agreed that the tool enables the accurate and appropriate impact analysis of changes in test cases and recommends the accurate and appropriate results based on this impact analysis (Q1 and Q2 respectively). Furthermore, all specialists strongly agreed (Q3) that they are confident in the high accuracy level of test case generation provided by this tool. For future application (Q4), 75% of specialists agreed that the tool will be applied to other systems

TABLE XIV. THE RESULT OF IMPACT ANALYSIS

System	Type of atomic changes	# Test cases		Precision	Recall	F-measure
		identified by an expert	identified by the tool			
KFD	Change of variable name	6	6	100%	100%	100%
	Change of variable type	35	35	100%	100%	100%
	Change of variable range value	-	-	-	-	-
	Change of number of variables	144	144	100%	100%	100%
	Change of variable order	185	185	100%	100%	100%
ONA	Change of variable name	21	21	100%	100%	100%
	Change of variable type	9	9	100%	100%	100%
	Change of variable range value	26,244	8,748	100%	33.33%	50%
	Change of number of variables	294	294	100%	100%	100%
	Change of variable order	26,568	9,072	100%	34.15%	52%

TABLE XV. SUMMARY OF THE IMPACT ANALYSIS OF TEST CASE BASE ON REQUIREMENT CHANGING

System	#Test Cases	Impact	Type of Changes (Number of Test Cases)								
			Variable Name	Data Type	Number of Variable		Variable Value			Variable Order	
					Add	Delete	Add	Delete	Update		
KFD	185	No Change	6	-	-	-	-	-	-	-	-
		Update	-	35	-	-	-	-	-	-	185
		Delete	-	-	27	-	-	-	-	-	-
		New	-	-	144	-	-	-	-	-	-
ONA	26,568	No Change	21	-	-	-	4,374	-	-	-	-
		Update	-	9	-	-	4,374	-	-	-	26,568
		Delete	-	-	-	2,592	-	-	-	-	-
		New	-	-	6	288	17,496	-	-	-	-

in the future. Moreover, specialists also gave feedback in the questionnaire (Q5). Most specialists said that the tool is accurate, reliable and easy to use. The tool enables the benefit in which the test case from the previous can be reused. However, some of them gave very useful feedback for the future improvement. They suggested that the tool should support other input file formats rather than just XML file format. We will use all feedback and suggestions from the specialists to improve the tool in the future. Overall, the specialists were mostly satisfied with our impact analysis tool.

TABLE XVI. SATISFACTION QUESTIONS

Questions	Average
Q1. The tool provides accurate and appropriate analysis of the impact of changes in test cases.	Likert scale (Mandatory)
Q2. The tool can offer accurate and appropriate results and recommendations based on the impact analysis.	Likert scale (Mandatory)
Q3. The tool can generate accurate and appropriate test cases.	Likert scale (Mandatory)
Q4. The tool can be applied to other system case studies in the future.	Likert scale (Mandatory)
Q5. Comments and suggestions for the tool application and improvement	Open-ended question

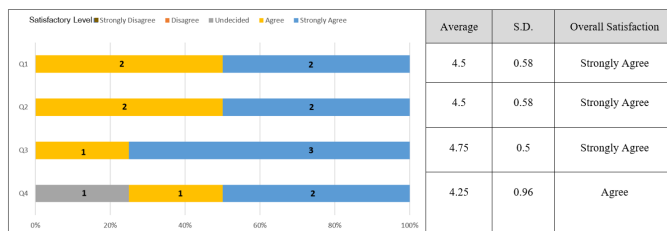


Fig. 14. Results of the four Likert scale questions

## V. LESSON LEARNED AND DISCUSSION

### A. Discussion

The key findings discovered from our developed tool is the ability to analyse the impact of test cases based on changes of use case based requirement specification that conforms to the proposed approach and framework. Based on the analysis and evaluation results, the developed tool can perform the impact analysis from the change of the requirements in two versions as before-change and after-change of the use case specification model according to the proposed five patterns of the cause of changes. These patterns are designed specifically for use case based specification including 1) change of variable name, 2) change of variable type, 3) change of variable range value, 4) change of number of variables and 5) changes of variable order. Compared with the study work of [8], they classified the cause of changes into seven patterns in which the first five patterns are the same as our proposed patterns. The other two patterns were not included as they are related to the change on web tag a link that is specific to web application. The decision after the impact analysis is also classified by this tool into four groups as *no-change*, *update*, *delete* and *create* groups as we expected.

Furthermore, CCTM is integrated into this tool to support automatic generation when new test cases require to be created. The validation and evaluation results confirm the key findings that our proposed approach supports all cases for merging two classification trees proposed by [19] as explained in Section IV-B and IV-C. From the validation and evaluation results, we can confirm that CCTM influences the benefit that reduces the time and increases the testing coverage for new test case generation. However, it has the limitation of testing a large and complex system that may not complete with a single classification tree as found in [6], [7]. We suggest that it needs to separate the testing into several testing units. From our experiences, for example, we separated the testing of our SUT,

MRS, into two testing units, one is for KFD subsystem and the other for ONA subsystem.

### B. Comparison of the Proposed Approach and Other Approaches

This section describes the comparison of our proposed approach with other approaches by [11], [9], [8]. Five criteria are considered for this comparison including *Impact Analysis Methods, System Domain, Source Type, Test Case Generation Techniques* and *Reusability Supports* as shown in Table XVII.

Our approach performs the impact analysis from the change of the requirements and classifies the cause of the change into five patterns as mentioned earlier. Comparing this with the approach proposed by [8], the cause of changes of a web application is classified into seven patterns in which the first five patterns are the same as our proposed patterns. Unlike the work of [11] and [9], it does not classify the cause of change into patterns during impact analysis. The existing test cases affected from the change is only analysed whether they can be reused.

Furthermore, as new test cases are required to be generated in the case that new features or requirements are added, CCTM, the hybrid testing approach that integrates ECP with CTM test case generation technique, is integrated into this tool to support automatic generation when new test cases require to be created. With the CTM technique, classification trees merging approach enables the test case redundant from two classification trees to be eliminated. Moreover, ECP technique in the CCTM provides the test case coverage of all possible scenarios as well as failure scenario. Comparing our implemented test case generation technique with the work of [8], [11] and [9]. In the work proposed by [8], their impact analysis framework provides ECP and BVA. These testing techniques only provide the test case coverage of all possible scenarios from the various range of separated partition. Unlike the work proposed by [11] and [9], they only focus on the impact analysis methodology that results in only the consideration of the affected test cases in the old version to be usable or unusable. Their framework, therefore, does not provide test case generation technique for the new test case generation.

Lastly, considering the comparison in terms of the reusability supports as demonstrated in Table XVII, our proposed approach provides the reusability level of test cases as a percentage from the impacted test cases that are implemented in a similar way with the work proposed by [8]. In the work proposed by [11] and [9], they classified only two levels of reusability of the existing test cases that are reusable or unusable.

## VI. CONCLUSION AND FUTURE WORK

This paper has demonstrated an automatic impact analysis approach of test cases based on changes of use case based requirement specification. Our experiment results with two case studies, KFD and ONA, have shown that the developed tool enables the benefits in which the impact on changes of test cases is analysed from the change of variables in use case specification. Two versions as before-change and after-change of the use case model are compared for consistency checking to detect the change. This results in the cause of variable

changes to be classified into five patterns of atomic changes are encoded in this tool. These classified patterns enable the impact on changes of the existing test cases to be analysed whether the existing test cases to be completely reused, partly updated as well as additionally generated. Consequently, the level of reusability of existing test cases is measured and the time to create the whole new cases for the after-changed version is reduced. Furthermore, CCTM, the hybrid test case generation technique encoded in the tool for generating new test cases influences the benefit to increase the level of testing coverage with a minimised number of test cases and reduces the redundant test cases as demonstrated. The beneficial contribution delivered by our proposed approach and tool is also confirmed by the validation and evaluation results from the practical specialists which are consistency with the results discovered from the researchers' perspective.

As our proposed approach and developed CIA tool supports the impact analysis of change of variable types that are only primitive programming data types including integer, floating-point number, boolean, character and string, the adjustment of the tool to support the ready-made "real world" data type e.g. date and time have been considered for the future work. To increase the capability and reliability of the tool, another research issue is the further evaluation of the tool with different system domains as suggested by the practical specialists.

## DEPLOYMENT AND AVAILABILITY

Our developed tool is available at <https://sites.google.com/phuket.psu.ac.th/testciatool/>. User guide manual document and source of example case studies (KFD and ONA subsystems) are also available on the website.

## REFERENCES

- [1] P. Jorgensen, *Software Testing: A Craftsman's Approach*, 3rd ed. Boca Raton, NY: Auerbach Publications, 5 2013.
- [2] I. Dewi, Y. Miftahuddin, M. Fattah, C. Palenda, and S. Erawan, "Point of sales system in inhome café website using agile methodology," *Journal of Innovation and Community Engagement*, vol. 1, no. 1, pp. 01–19, Mar 2021.
- [3] S. Purwanti, A. Febriani, M. Mardeni, and Y. Irawan, "Temperature monitoring system for egg incubators using raspberry pi3 based on internet of things (iot)," *Journal of Robotics and Control (JRC)*, vol. 2, no. 5, pp. 349–352, 2021.
- [4] I. R. Munthe, B. H. Rambe, R. Pane, D. Irmayani, and M. Nasution, "Uml modeling and black box testing methods in the school payment information system," *Jurnal Mantik*, vol. 4, no. 3, pp. 1634–1640, 2020.
- [5] S. Sutiah and S. Supriyono, "Software testing on e-learning madrasahs using blackbox testing," *IOP Conference Series: Materials Science and Engineering*, vol. 1073, no. 1, p. 012065, Feb 2021.
- [6] M. Grochtmann and K. Grimm, "Classification trees for partition testing," *Software Testing, Verification and Reliability*, vol. 3, no. 2, pp. 63–82, 1993.
- [7] M. Grochtmann, K. Grimm, J. Wegener, and D.-b. Ag, "Tool-supported test case design for black-box testing by means of the classification-tree editor," in *Proceedings of the EuroSTAR (EuroSTAR, 1993)*, 1993.
- [8] S. Phetmanee and T. Suwannasart, "A tool for impact analysis of test cases based on changes of a web application," in *Proceedings of the International MultiConference of Engineers and Computer Scientists 2015*, 2015, pp. 497–500.
- [9] T. Sakkarinkul and T. Suwannasart, "Test case impact analysis from use case description changes," in *Proceedings of the International MultiConference of Engineers and Computer Scientists 2015*, 2015, pp. 523–527.



TABLE XVII. COMPARISON OF FEATURES AND CAPABILITIES OF OUR PROPOSED APPROACH WITH COMPARABLE WORK

Features	Proposed Approach	[8]	[11]	[9]
Impact Analysis Methods	Indicate and classify the impact analysis result into five patterns	Use seven patterns of the change of a web application to classify the impact analysis result	Analyse the impact from the change list of use case specification but do not classify the pattern of changes	Analyse the impact from the change list of use case specification but do not classify the pattern of changes
System Domain	Hospital Information System	Web Application	Web Application	Web Application
Source Type	Use Case Spec.	GUI spec.	Use Case Spec.	Use Case Spec.
Test Case Generation Techniques	Use the CCTM (ECP and CTM) technique	Use ECP and BVA techniques	No test case generation	No test case generation
Reusability Supports	Calculate the percentage of the reusability level of test cases	Calculate the percentage of the reusability level of test cases	Only indicate that the impacted test cases can be reused or updated	Only indicate that the impacted test cases can be usable or unusable

- [10] A. Intana and T. Sriraksa, "Impact analysis framework of test cases based on changes of use case based requirements," in *Proceedings of the 23rd International Computer Science and Engineering Conference (ICSEC, 2019)*. IEEE, 2019, pp. 230–235.
- [11] M. Raengkla and T. Suwannasart, "A test case selection from using use case description changes," *Lecture Notes in Engineering and Computer Science*, vol. 2202, pp. 507–510, Mar 2013.
- [12] C. Sriarpanon and T. Suwannasart, "A source code and test cases impact analysis tool for database schema changes," *Lecture Notes in Engineering and Computer Science*, vol. 1, pp. 466–469, Mar 2015.
- [13] A. Kampeera and T. Suwannasart, "Impact analysis to database schema and test cases from inputs of functional requirements changes," in *Proceedings of the International MultiConference of Engineers and Computer Scientists 2016*, Mar 2016, pp. 449–453.
- [14] N. Cherdsakulwong and T. Suwannasart, "Impact analysis of test cases for changing inputs or outputs of functional requirements," in *2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. IEEE, 2019, pp. 179–183.
- [15] WHO, "Mean body mass index," <https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi>, Jun 2021.
- [16] B. Ramadoss and P. Prema, "An approach for merging two classification-trees," in *Proceedings of the IEEE International Advance Computing Conference 2009*, 2009, pp. 1602–1607.
- [17] K. M. Ting, *Precision and Recall*. Boston, MA: Springer US, 2010, pp. 781–781. [Online]. Available: [https://doi.org/10.1007/978-0-387-30164-8\\_652](https://doi.org/10.1007/978-0-387-30164-8_652)
- [18] A. L. Imoize, D. Idowu, and T. Bolaji, "A brief overview of software reuse and metrics in software engineering," *World Scientific News*, no. 122122, p. 56–70, 2019.
- [19] B. Ramadoss, P. Prema, and S. R. Balasundaram, "Combined classification tree method for test suite reduction," in *Proceedings on International Conference and workshop on Emerging Trends in Technology (ICWET, 2011)*, no. 11, 2011, pp. 27–33.

# Trust Management for Deep Autoencoder based Anomaly Detection in Social IoT

Rashmi M R<sup>1</sup>, C Vidya Raj<sup>2</sup>  
Research Scholar, VTU, Belagavi<sup>1</sup>

Department of Computer Science and Engineering, NIE, Mysore, India<sup>1,2</sup>

**Abstract**—Social IoT has gained huge traction with the advent of 5G and beyond communication. In this connected world of devices, the trust management is crucial for protecting the data. There are many attacks, while DDOS is the most prevalent BotNet attack. The infected devices earnestly require anomaly detection to learn and curb the malwares soon. This paper considers 9 IoT devices deployed in a Social IoT environment. We introduce a couple of attacks like Bash lite and Mirai by compromising a network node. We then look for traces of malicious behavior using AI algorithms. The investigation starts from a simple network approach - Multi-Layer Perceptron (MLP) then proceeds to ML - Random Forest (RF). While MLP detected the malicious node with an accuracy of 89.39%, RF proved 90.0% accurate. Motivated by the results, the Deep learning approach - Deep autoencoder was employed and found to be more accurate than MLP and RF. The results are encouraging and verified for scalability, efficiency, and reliability.

**Keywords**—Social IoT; trust management; anomaly detection; DDoS; deep autoencoder

## I. INTRODUCTION

IoT is a disruptive network technology that has advanced quickly over the past ten years in every technology field, including smart cities, satellites, smart homes, smart businesses, smart transportation, and smart healthcare [1]-[4]. It consists of several IoT devices (Things) that may gather and share data through the conventional internet thanks to their various sensors, actuators, storage, computing, and communication capabilities [5]. The industry's security concerns resulting from the enormous range of IoT devices and vendors. On the off chance that security and protection are not accommodated their organizations and information, partners are probably not going to broadly embrace IoT innovations. Recent cyber security reports [6] have revealed that assaults against IoT settings have increased in frequency due to the IoT ecosystem's expanding attack surface, which extends from the edge to the cloud [7]. Therefore, a significant ongoing problem for engineers in this industry is designing and creating secure IoT systems [8]. The sensitive nature of the data collected and processed within the IoT network necessitates security from potential breaches. As the first line of defense against potential security attacks [9]-[10] on weak devices [11], like distributed denial of service (DDoS) attacks [12], various security mechanisms are currently used to protect sensitive data. These mechanisms include firewalls, authentication protocols, encryption methods, antivirus software, and more. Such assaults are carried out against another network entity, such as a business or a government, by a collection of infected machines (bots) that are part of a botnet and are under the attacker's control via a C&C (Command and Control) server. Due to the extensive

use of data, several new anomalies—original and mutations of previously observed anomalies—are often produced.

The Compromise of IoT devices and their enrollment into IoT botnets under attackers' control is one of the main threats to IoT networks and devices. Well-known IoT botnets like BASHLITE and Mirai continue to pose substantial DDoS risks, according to the conclusions of a Hundred of active command and control (C & C) servers are included in the Distributed Denial of Service (DDoS) report for the first quarter of 2021 [13]. Due to the IoT's primary characteristics, which must be considered: heterogeneity, scalability, and limited resources, mitigating such threats could be very difficult (power, memory, and processor). As a result, creating solutions for IoT environments that can detect aberrant behaviors and assaults has emerged as a major problem in the field of IoT cyber security and a hot topic for researchers. Subsequently, an IoT organization can profit from extra protection from security attacks thanks to an anomaly identification framework that can act as a second line of guard. Furthermore, enterprises in this market sector are primarily focused on fusing IoT technology with other slashing technologies like AI (AI algorithms are used for data processing and analysis), Big Data (handling a huge amount of information from IoT devices), or 5G connectivity (mobility and broadband links for IoT sensors). Since it merges AI with IoT, the AI of Things (A-IoT), a disruptive technology that aims to analyze data to make autonomous and automated decisions on IoT networks, is receiving special attention [14].

To increase anomaly detection accuracy, researchers have recently looked into machine learning (ML) and deep learning (DL) techniques. Studies have shown that both ML and DL approaches are useful for extracting characteristics from network traffic that can be used to classify the traffic as benign or abnormal [15]. The DL has shown effective at learning relevant attributes from the raw data because of its deep model, which provides a variety of abstractions for learning intricate features for precise predictions [16]. Due to the huge volume of data generated by IoT devices, these characteristics of DNN have made it a suitable methodology to be adopted for anomaly detection schemes created for IoT networks [17]. Larger networks, however, might find these solutions impractical. In addition, because a separate model needs to be updated and maintained for individual IoT sensor, using several auto encoders might make network security challenging to establish and administer. We offer three methods. Deep auto-encoder, Random Forest, and Multilevel Perceptron for identifying anomalies in IoT networks by observing and analysing the innocuous "snapshots" of behaviour from each IoT device, we suggest a deep auto-encoder-based anomaly

## II. BACKGROUND

### A. Categorization of IOT Anomalies

An anomaly is a data point in a modelled system that is not aligned to the normal behaviour. Rare occurrences or observations known as anomalies differ dramatically from typical behaviour or patterns seen in a single data point or throughout the full dataset. An aim of algorithm will probably find an irregularity's event and characterize/gather its objective in light of the fact that in principle, anomalies are brought about by outside powers like sensor breakdown or outer assault. The estimation structure that best matches the expected information conduct is fundamental in the twofold order of a peculiarity. Additionally, each application needs a unique detection approach due to the complexity of numerous circumstances [19].

Taking into account the arrangements from earlier examinations like Fahim and Sillitti [20] and Cook et al. [21], an IoT peculiarity discovery approach is partitioned into four classes. According to the problem they address, how they are used, the kind of method used, and the algorithm's latency, they are divided into different categories. Fig. 1 presents an example overview of the four groups.

(i) **By Method:** The methods can use machine learning, statistics, or geometrical methods. Geometrical approaches are based on the presumption that the anticipated and abnormal data are separated when distance- and density-based representations of a given dataset are used. The reasoning behind detachment or density-based approaches is that peculiarities arise in scanty districts in a bunch of data of interest. These procedures classify irregularities utilizing a static or dynamic magnitude value ( $t$ ) on the assessed distance ( $d$ ), which is given as:

$$d = \begin{cases} < t, Normal(underthreshold) \\ > t, Anomaly(abovehethreshold) \end{cases} \quad (1)$$

(ii) **By Application:** The three methods that an application uses to categorize anomalies are data cleansing, data destruction, and constructive categorization. The world of elderly people's everyday activities to ensure safe and its evaluation of the efficiency of multilayer perceptron (MLP), SVM classifiers, and k-nearest neighbours (KNN), offer value when applications are constructive or positive in nature.

(iii) **By Anomaly Type:** The situation-specific type, such as the point, contextual, and collective, is one of the most frequently observed types. If just one piece of data deviates from the norm, it qualifies as an anomaly. The identification of fraud with credit cards is one instance. An incident that might be seen as abnormal in a certain context is called a contextual anomaly. The last kind of anomaly, collective anomaly, examines the complete dataset, unlike a point or contextual anomaly.

(iv) **By Latency:** Whether a detection technique is conducted immediately during the data collecting stage or after it has been stored depends on its latency and scalability [18]. An online technique can serially analyse data with a single point of information or window without having access to the

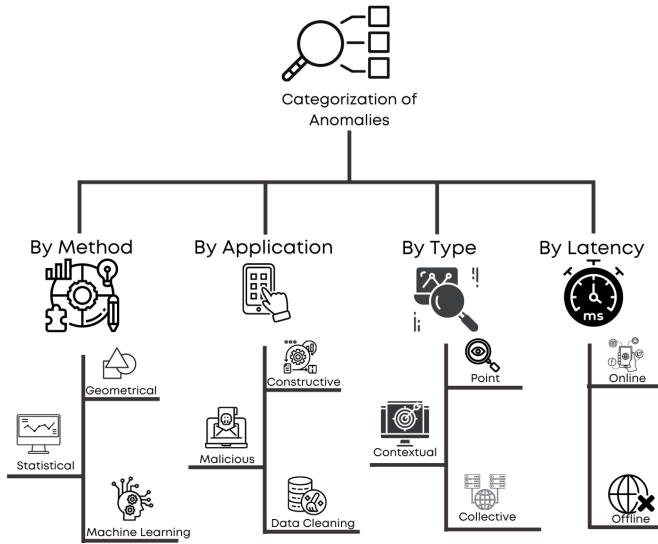


Fig. 1. Categories of anomaly in IoT

detection system as a viable method for identifying botnet attacks. The experiment is built on a test bed network of nine IoT devices and simulates the BASHLITE and Mirai botnets, two well-known botnets. Auto-encoders well defined for each IoT gadget is used to gain proficiency with the regular traffic properties and to caution when they can't recreate the harmless traffic samples.

The main contribution of the work is three-fold:

- First, a brief review of the constraints and vulnerabilities of Social IoT networks, various attacks, and anomalies is presented.
- Next, we deploy 9 IoT devices in a Social set-up while injecting one node with BASHLITE and Mirai infection resulting in the DDOS attack.
- Finally, we employ ML, NN, and DL-based approaches - RF, MLP, and Deep Auto-encoder- to detect anomalies accurately and ascertain the results for scalability, efficiency, and reliability.

The following is how the paper is structured. Section II introduces anomaly detection and discusses numerous sorts of anomalies in the IoT context. Furthermore, it situates the topic within the framework of DDoS attacks and gives a taxonomy of DDoS attacks. Section III discusses anomaly detection strategies and forms of anomaly attacks in the social IoT and finishes with the deep autoencoder as a solution for anomaly identification. The proposed framework for anomaly detection and the evaluation metric to validate the performance of the proposed method are then presented (Section IV). We then examine the intriguing results that demonstrate the superiority of the suggested strategy over rival schemes in Section V. The final remarks are discussed in Section VI.

TABLE I. ATTACK TYPE CLASSIFICATION FOR IoT

Sl.No	IoT Attack	Description
1	Dos	This type of attack involves the deliberate sending of bugs or packets to render resources unavailable to hosts connected to the internet.
2	Data Type probing	This exploit involves a hostile attacker writing an unintended data type .
3	Malicious Control	This type of attack allows the attacker unapproved access to the user's system
4	Malicious Operation	In general, the malware was a factor in this attack. Attacker engages in bogus activity on a system that has been authenticated
5	Scan (SC)	In this assault, equipment sporadically scans the framework for information to obtain, which might ruin information
6	Spying (SP)	In this attack, the attacker targets the system's weak spots and gains access via a backdoor to steal sensitive data
7	Wrong Setup	It is an assault, wherein culprit deliberately sends bugs or parcels for making asset inaccessible for the host associated with the web
8	Normal	This attack may have been planned or unintentional, but it might still cause harm by upsetting the system.

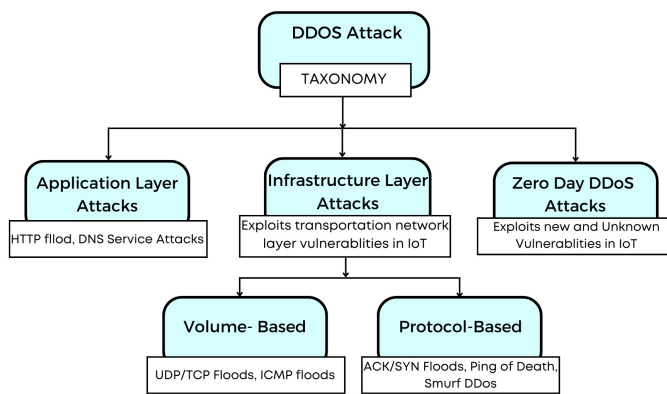


Fig. 2. Taxonomy of DDoS attacks

complete input. Traditional and online geometry and statistical methodologies include the distance-, density-, and angle-based approaches listed above. Offline algorithms, however, have complete access to the information. They employ difficult, computationally expensive, sophisticated methods to solve the problem in a reasonable amount of time. Anomalies are brought on by external forces like sensor malfunction or an external attack. The various types of attacks in IoT have been summarized in Table I. The most common kind of assault that may be launched against any application is a DDoS attack. DDoS attacks are the loudest kind of cyber attacks.

### B. Distributed Denial of Service Attacks

The DDoS attack is undertaken to overwhelm the target and interrupt services, as the name suggests. IoT devices are highly suited for the DDoS attack because it needs a lot of devices to conduct an attack. Users won't recognise that the gadget is compromised, as is typically the case. There is a pressing need to identify assaults quickly in order to remove affected devices as the number of IoT devices grows. BASHLITE and Mirai employed IoT devices as Botnets in a large DDoS attack, and other similar attacks have also occurred [22]. The various types of DDoS attacks have been illustrated in Fig. 2.

DDoS attacks have demonstrated a variety of attacking strategies over the years, and a variety of potential attacks are continuously being tested. IoT-specific DDoS attack strategies are not much different from conventional DDoS attack strategies. They use similar methods to take advantage of flaws in

both IoT devices and conventional systems. However, because of the heterogeneity present in IoT devices, DDoS assaults targeted specifically at IoT are more varied and complex. We will use the fundamental layered architecture of an IoT network to categorise DDoS assaults in this section.

Three different types of DDoS attacks are illustrated in a comparative analysis in Table II. DDoS assaults have affected well-known service companies like Amazon Web Services. AWS, Cloud are, KrebsOnSecurity, and other security service providers against similar assaults are also DDoS attack victims. Therefore, assaults on these significant institutions affect enterprises financially and reputationally. In a DoS attack, the attacker makes bogus requests using the target's resources in an effort to disrupt the target's services. DDoS involves simultaneous demands coming from several sources. DDoS attack mitigation becomes challenging as a result. There are many different types of DDoS attacks, such as Teardrop, Smurf, TCP SYN Flood, Smurf, Teardrop, Botnet attack and Ping of Death. DDoS assaults can also be categorised as amplification and reflection assaults. The request and response sizes are equal in a reflection attack [23], however in an amplification assault, the response size is significantly larger than the request size [24].

1) *Compromising an IoT device (BoT) and BotNets:*  
Due to the inherent characteristics of botnets, namely the existence of widely dispersed peers and C&C servers across the Internet with masked communication techniques, there is no secure strategy that can be utilized to shut down all bot movement without disrupting real traffic. Bot malware like Mirai actively searches the network for weak points, hunting for devices that allow unauthenticated access or that use weak or default credentials. After the defence is broken, a concise bootstrap script is executed, which downloads the whole program from the C&C. Other methods of spreading the dangerous bot code include the widespread use of phishing emails and freeware promotions to trick people into downloading it on their PCs. Making sure the bot binaries avoid antivirus programmes, which often employ signature-based detection techniques, is just as crucial as the bot binaries' distribution mechanism. It was found that Storm was doing this by repeating the encoding its un authorized two times every hour. Although, as IoT sensors lack the processing capacity required to run sophisticated anti-virus software, botnets that target them may conveniently ignore this complexity.

The rallying phase, which occurs after infection, entails

TABLE II. ATTACK TYPE CLASSIFICATION FOR IOT

Variety of Attack	Attacker's Objective	Size Measured in	Examples	Existing Countermeasure to the Attack
Volumetric Attack	Take up the entire bandwidth between the target and the internet	Bits per Second (bps)	UDP, TCP, DNS floods and amplification of NTP	On-demand scalable scrubbing centres get the rerouted traffic so they can handle it.
Protocol Based Attack	To use server, firewall, and load balancer resources	Packets per Second (pps)	Flooding the TCP SYN, Death Ping, Smurf Attack	Recognizable proof method is utilized regularly to separate among authentic and ill-conceived traffic to impede the assault prior to arriving at the objective server.
Application Layer Attack	To exhaust target resources	Requests per Second (rps)	HTTP Flood, DNS Flood	These are by and large sluggish assaults and relieved by recognizing bot conduct utilizing manual human tests and comparative procedures

developing a hidden method for receiving instructions from the CC [25]. This stage's fundamental objectives are to hide the address of the C&C and ensure that any orders shipped off the bots are encoded. Among the components are the "fast flux" strategy (Tempest), which rapidly pivots the C&C server's tends to behind a DNS name, and the utilization of Domain-Generation Algorithm (DGA) [26]-[27], Which require each recently tainted machine to attempt to determine haphazardly created area names to recognize its C&C. Later changes enjoy taken benefit of distributed correspondence, which further clouds the C&C [28]-[29].

### III. ANOMALY DETECTION TECHNIQUE

Anomaly or outlier detection problems can be used to frame the task of identifying an assault. This is predicated on the idea that malware and regular network traffic would differ in certain ways, allowing an algorithm to distinguish between the two. In this piece of research, we employ deep learning techniques. The utilization of sophisticated artificial neural networks architecture, that is modeled after the human brain and compute in a completely different way from conventional digital methods, is the cornerstone of deep learning approaches. To learn the weights of the network and create a model that can distinguish between assaults and normal behaviour, deep networks in a NADS technique need to know something about the valid data class.

The utilization of autoencoders in non-linear cooperation and for applications requiring network traffic highlights is an expected system for anomaly identification. Autoencoders can learn more effectively with less training data when depth is used because it lowers the computational cost of modelling functions [30]. These affirmations propelled us to test a Deep autoencoder model for irregularity detection in IoT system. An autoencoder is a neural network-based unsupervised learning model that has been trained to reconstruct the input into the output. It is made up of two parts: an encoder and a decoder. The encoder is used for input, and the decoder is used for output (code).

#### A. Deep Autoencoder

Deep Auto Encoder (DAE) is a tool for unsupervised learning of effective coding. The simplest DAE architecture consists of an input layer, several hidden layers, and an output layer that contains the same number of neurons for reconstruction as the input layer. It becomes a deep autoencoder when both the encoder and the decoder, the two parts of the autoencoder, are deep networks. The decoder's layers are inverted, but

both devices share a similar construction. A deep autoencoder features of Deep autoencoder are:

- 1) A array  $X$  representing  $n$  dimension input data, where  $X = (X_1, X_2, \dots, X_n)$
- 2) Fig. 3 shows several hidden layers that stand in for various encoding and decoding levels. These layers produce an irregular illustration of the data input and reconstruct it for the output layer.
- 3) The array  $X' = (X'_1, X'_2, \dots, X'_n)$  is an output layer. The output, which is a recovered copy of the input data, is the same size as the input.

4) In addition to weights and biases, an activation function. An activation function is used by each neuron in a layer to determine its output based on the weighted sum of its input.

There are two types of activation functions utilised in DL models: direct capabilities and non-straight capabilities. The most famous nonlinear actuation capabilities are the sigmoid (calculated), exaggerated digression, and corrected direct unit (ReLU). The exaggerated digression capability is generally used in two-class characterization, the sigmoid capability is explicitly utilized while determining the result as a likelihood, and the ReLU capability is the most often utilized in basically all profound brain organizations. In our examination, we applied the Sigmoid capability (1) to the last layer of the decoder and the ReLU capability (2) to each secret layer of the autoencoder.

$$S(x) = \frac{1}{1 + e^x} \tag{2}$$

$$R(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{3}$$

A machine learning model's learnable parameters are weights and biases. The biases and weights are assigned to the inputs before they are passed across neurons. While biases guarantee that neuron activation will still occur even if all the inputs are zeros, loads show the amount of impact the info possesses on the result. In a profound autoencoder, layer  $l$  is addressed by  $W_{ij}^l$ , which addresses the weight applied to the connection between hub  $j$  of layer  $l - 1$  and hub of layer  $l$ , and  $b_i^{(l)}$ , which addresses the predisposition connected with the hub. Following formula is used to determine neuron  $i$ 's output value from layer  $l$ .

$$O_i^l = F\left(\sum(W_{ik}^{(l)}) + b_i^l\right) \tag{4}$$

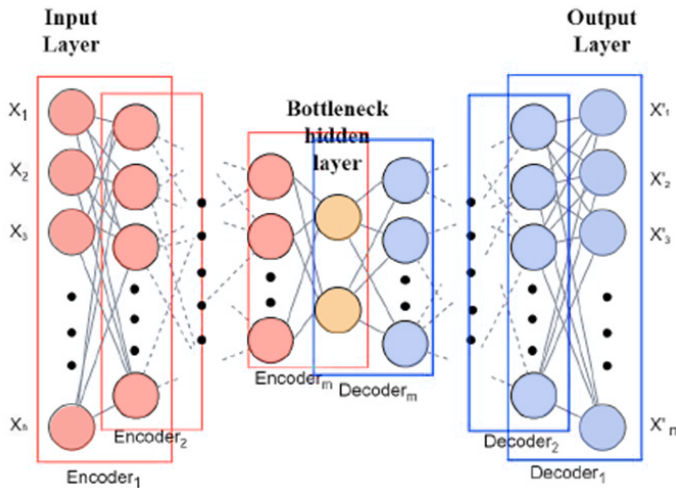


Fig. 3. Deep autoencoder

where  $F$  denoted as activation function,  $x_j$  is a neuron's input value that was acquired from the layer  $l - 1$  output of node  $j$ . Nodes on the subsequent layer  $l$  output from node  $l$  as input. This input data is often transformed into an encoding map by autoencoders, which is then further decoded to produce an output layer that represents the input layer's recovered version.

$$X' = D(E(X)) \quad (5)$$

To utilize autoencoders, E and D should be prepared to lessen the contrast among  $X$  and  $X_0$ . An expense capability that works out the mistake between the real and expected values is utilized to contrast the delivered yield  $X_0$  with the info  $X$  (that is supposed to be created). The model loads are changed during preparing until a decent planning of contributions to yields is delivered to lessen blunder (cost). Deep autoencoders can also use the ANN-specific loss functions. The Mean Squared Error (MSE) is frequently employed. Binary cross-entropy loss is recommended if the input solely contains binary values.

#### IV. PROPOSED METHOD

The research work's methodology is presented in this part. Everyone is aware that the Internet of Things (IoT) is susceptible to a wide range of assaults, including network, software, physical, and privacy-related ones. The new safe IoT framework we provide here allows for the detection of attacks in the IoT environment using a Deep Auto-encoder method. Information extraction from IoT sensor organizations, information preprocessing, information cleaning, highlight extraction, preparing profound learning models, irregularity identification, and effectiveness estimation with accuracy, confusion matrix, recall, and FPR and TPR curves are just a few of the processes that are integrated into the overall framework. This methodology works well for detecting attacks and anomalies in IoT infrastructure. Fig. 4 shows the proposed framework's overall image, which combines numerous distinct sub-processes.

#### A. Data Source

Our research's main objective is to create a clever, secure, and trustworthy framework for identifying anomalies and assaults in IoT sensor networks. The N-BaIoT dataset, an open-source dataset obtained from Kaggle, is used for experiments with our model. The accuracy, recall, and confusion matrices of the dataset are used to assess the model's efficacy. Our experiments' foundational dataset, N-BaIoT, has likewise been utilized in various examinations on botnet assault location. The vast majority of them utilize parallel or multi-class characterization and classification-based techniques. Our goal is to find an appropriate approach to identifying IoT network traffic anomalies without labelling the raw data first. Network sniffing tools are used to intercept IoT network communications. Some freely accessible tools, such tcp dump and Wireshark, can be used for this. The primary duty is to record network packs, which must subsequently be examined and visualised for analysis. A dataset contains the features that have been retrieved from the network packets.

#### B. Data Processing

1) *Data pre-processing*: Data pre-processing is an important stage in learning theories since network data derived from network activity also include these data, which are typically loosely regulated and lead to irrelevant or redundant data values. It cleans up network data by removing unnecessary, distracting, or irrelevant information, which enhances the effectiveness of DE techniques for identifying attack behaviours. The following describes the production, reduction, conversion, and normalisation of features as part of data pre-processing for network data. The preprocessed separated highlights are then used to wipe out repetitive streams, standardize consistent elements, and onehot encode straight out highlights.

2) *Data cleaning*: We initially locate and eliminate mistakes and duplicate values from the dataset in this process. After that, enter a specific value as "NaN" to replace any missing values. Any machine learning algorithm or model's accuracy and effectiveness can be improved with the aid of this procedure.

3) *Feature extraction*: We take a conduct depiction of the hosts and conventions that imparted this parcel each time a bundle shows up. The depiction accumulates traffic measurements over various fleeting windows to order the information that was all sent between the source and objective IPs (channel), the source and objective Macintosh addresses, the source and objective TCP/UDP attachments, and the source and objective IPs overall (attachment). Similar arrangement of elements are removed across different time spans. These attributes may be quickly and incrementally calculated, making it easier to identify fraudulent packets in real time. Additionally, despite being general, these qualities can catch specific activities such source IP ridiculing [2], a component of attacks by Mirai. For example, the highlights collected by the Source MACIP, Source IP, and Channel will rapidly uncover a critical irregularity inferable from the concealed conduct coming from the faked IP address when a compromised IoT gadget parodies an IP.

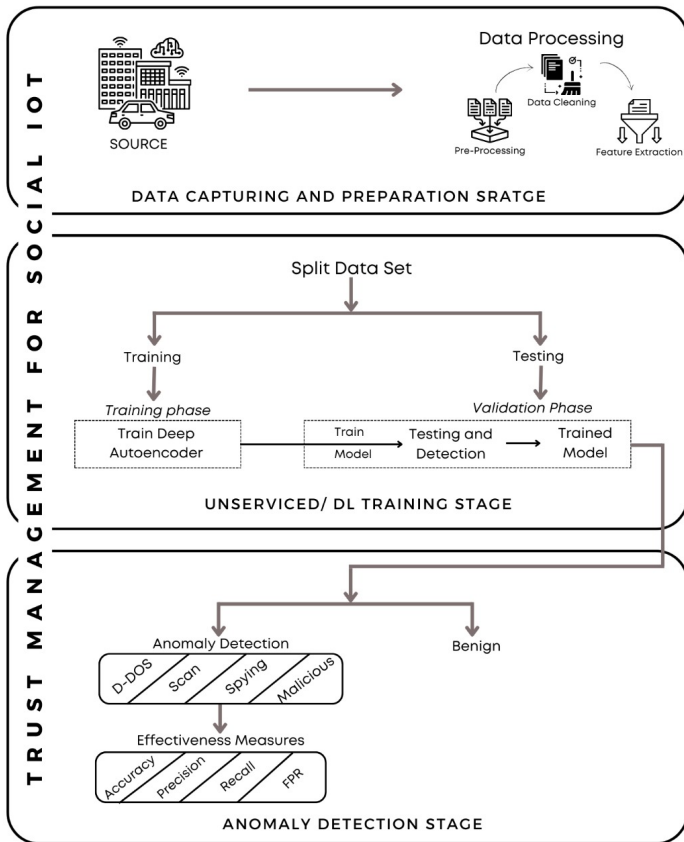


Fig. 4. Proposed Framework for Anomaly Detection

### C. Split Dataset

- **Training Data:** We prepared and streamline a profound autoencoder on 2/3 of the harmless information from every one of the nine IoT gadgets (i.e., the preparation set of every gadget). To record regular organization traffic designs, this was finished.

- **Testing Data:** Every one of the malevolent information as well as the excess third of harmless information made up every gadget's test information. We utilized the relating trained autoencoder as an anomaly finder on each test set. The detection of anomalies—the hacks perpetrated from each of the aforementioned IoT devices—was successful 100% of the time.

### D. Train the Deep Autoencoder

The Harmless and Inconsistency samples are then picked as the objective highlights for the DNN's training on the Train dataset utilizing twofold characterization. A prepared DNN model is made after this stage. We utilize Deep autoencoder as our essential anomaly identifier and exclusively keep a model for each IoT gadget. A neural network that has been trained to adapt its contributions after some upgradation is called an autoencoder. The system will become familiar with the connections between its feedback highlights and important ideas on account of the pressure. An autoencoder will find lasting success at reproducing typical perceptions on the off chance that it is exclusively prepared on harmless cases, yet it

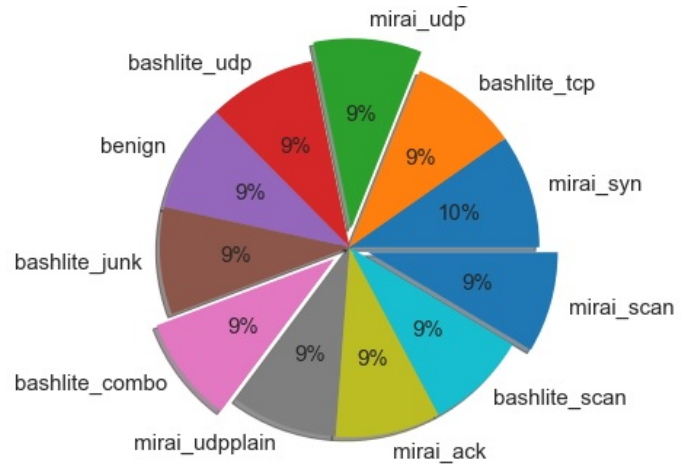


Fig. 5. Distribution of benign and infected devices in dataset

will neglect to recuperate unusual perceptions (obscure ideas). We mark the gave perceptions as strange when a sizable reproduction blunder is found.

We amplify the genuine positive rate (TPR, distinguishing assaults when they happen) and diminish the bogus positive rate in each preparing model by streamlining its boundaries and hyperparameters (FPR, wrongly stamping harmless information as vindictive). The model learns examples of ordinary movement from two different datasets that are used for training and optimization and only contain benign data.

### E. Testing / Detection

The Test dataset is then used to put the trained model to the test, identifying records as either benign or anomalous flows. Benign traffic was permitted to pass through unimpeded if it was anticipated. On the other side, if an anomaly is anticipated, the network administrator is alerted to take further action.

In order to categorise each instance as benign or anomalous, we eventually apply the improved model to feature vectors collected from constantly monitored packets. Then, whether the entire related stream is benign or anomalous is determined by a majority vote on a series of marked occurrences. As a result, if an abnormal stream is found, an alarm might be sent because it might be a sign of malicious activity on an IoT device.

### F. Evaluation Metric

It is possible to calculate accuracy using the proposed deep learning framework. Additionally, we can quantify the complexity of our deep learning models, which refers to how many parameters the model contains and how much weight it has when stored to disc, as well as how long the training process takes to obtain good accuracy in terms of time in seconds. The accuracy, precision, false-positive rate, and recall may then be determined for the task of anomaly detection system, which determines the proportion of actual properly recognises instances by model. Eq. (3) demonstrates how to compute recall using (8).

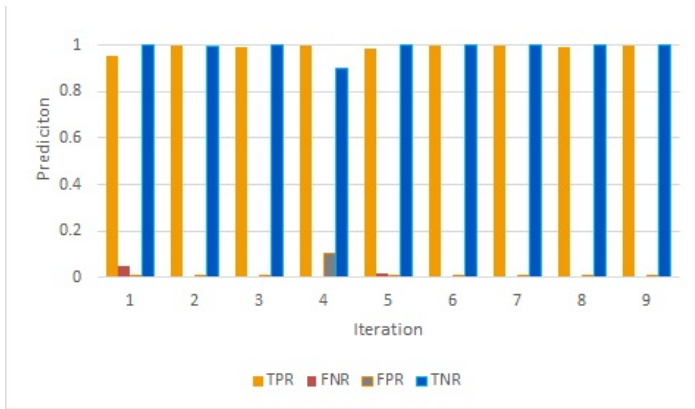


Fig. 6. Prediction probability factor for multi layer perceptron

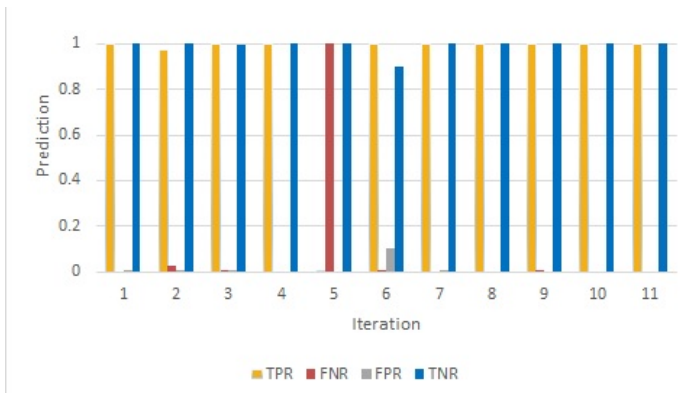


Fig. 7. Prediction probability factor for random forest

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$PR = \frac{TP}{FP + TN} \quad (7)$$

$$FPR = \frac{FP}{FP + TN} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

Whereas, TP termed as True Positive, FP as False Positive, TN as True Negative, and FN as False Negative, FPR as False Positive Rate, TPR as True Positive Rate, TNR as True Negative Rate, and FNR as False Negative Rate.

## V. RESULTS AND DISCUSSION

Python was utilised as a platform for implementing the suggested model, and experiments were carried out on an N-BaIoT dataset of an IoT sensor environment. The nine innocuous data sets that we gathered correlate to the nine IoT devices. Fig. 5 shows how different infected and healthy devices are distributed throughout the dataset.

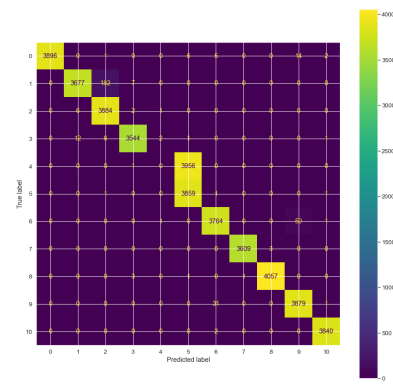


Fig. 8. Confusion matrix of multi layer perceptron

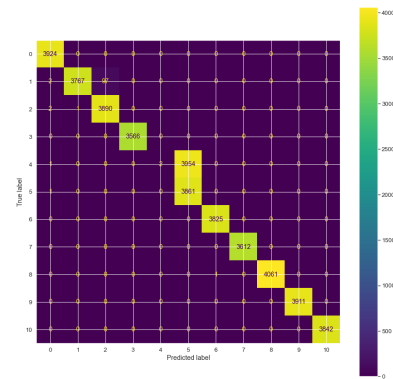


Fig. 9. Confusion matrix of random forest

We offer a comparison of the proposed model, MLP, and RF schemes, with the results for the performance metrics Accuracy, confusion matrix, Precision, F1-Score, Recall and TPR, TNR, FPR, FNR, to confirm the accuracy of the suggested technique. The results of the performance metric taken into account in this study with regard to MLP and RF schemes are summarised in Table III. It is easily shown that the multi-layer perceptron neural network scheme is outperformed by the machine learning-based random forest scheme.

The likelihood of prediction rate for TPR, FPR, TNR, and FNR in MLP and RF systems is shown in Fig. 6 and 7. Misunderstanding Matrix is also constructed for MLP and RF techniques, as shown in Fig. 8 and 9, respectively, to make it simple to spot class-related confusion. It is also referred to as an error matrix and is offered as a table matrix for displaying algorithmic performance and ambiguity in classifier predictions. The performance of the RF technique would not, however, provide the accuracy that is promised as the IoT network's size increases. In light of the massive data set generated by the IoT environment's many IoT devices, the Deep learning-based autoencoder approach is seen as a viable mechanism.

We choose the Window size as 82, Learning rate as 0.01, Optimizer as Adam, activation function as Relu in encoder, and Relu and Sigmoid are utilised in decoder while building the deep learning based autoencoder, referred to as deep autoencoder in this work. The training process's Loss function is the mean square error. Tensorflow was utilised for training.



TABLE III. PERFORMANCE SUMMARY OF MULTI LAYER PERCEPTRON AND RANDOM FOREST

Scheme	Accuracy	Recall	F1 Macro	F1 Micro	TPR	FNR	FPR	TNR
MLP	89.39%	89.52%	0.86	0.89	0.89386	0.16014	0.01061	0.98939
RF	90.0%	91.0%	0.87627	0.90482	0.90482	0.09518	0.00952	0.99048

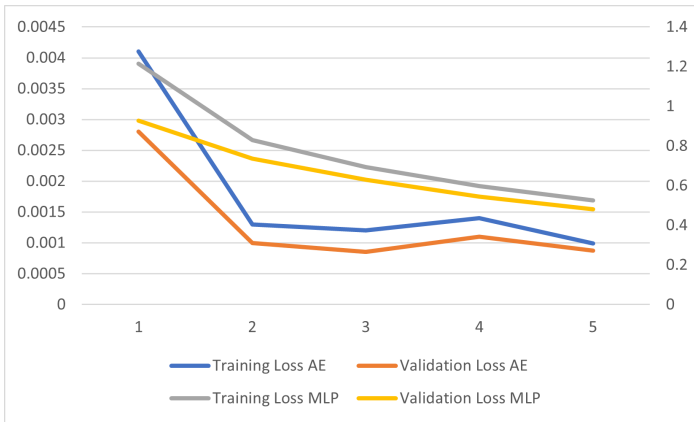


Fig. 10. Loss function of deep autoencoder

The dimension of the input layer for each autoencoder was equal to the number of features in the dataset (i.e. 115). For proper compression of the input layer between encoder and decoder and to reflect its fundamental properties, the autoencoder must effectively execute dimensionality reduction internally.

It is important to appraise the blunder for the model's present status as a component of the improvement technique more than once. To refresh the loads and lower the misfortune on the ensuing assessment, it is important to choose a mistake capability, otherwise called a misfortune capability, that might be utilized to gauge the deficiency of the model. A planning from contributions to yields is advanced by brain network models through models, and the misfortune capability utilized should be suitable for the particular prescient displaying task being tended to, like grouping or relapse. Also, the result layer's design should be reasonable for the chosen misfortune capability. Fig. 10 compares the loss functions for MLP and Deep autoencoder models; when the number of IoT devices is lower, both techniques perform similarly (up to 2). However, when the number of IoT devices grows, the suggested deep autoencoder model outperforms MLP in terms of performance.

The performance of the Deep autoencoder is summarised in Table IV. When compared to MLP and RF techniques, the suggested Deep autoencoder methodology performs better. The outcomes attest to the proposed IoT network solution's superiority. Additionally, it is noted that among the three schemes taken into account in this study, the RF technique is the second-best model and the MLP is the poorest.

## VI. CONCLUSION

Current internet security measures, such as firewalls and gateways, are ineffective at identifying sophisticated and unidentified assaults in an IoT environment. It is essential to secure this network infrastructure as demand for IoT networks grows. This study explains how AI works to identify assaults

TABLE IV. PERFORMANCE SUMMARY OF DEEP AUTOENCODER

Node Number	Shape of Data	Detected Anomalies
1	(22154, 115)	0.0%
2	(96781, 115)	100.0%
3	(60554, 115)	100.0%
4	(65746, 115)	100.0%
5	(156248, 115)	99.94%
6	(56681, 115)	100.0%

and anomalies in the environment of IoT sensors. The detection and classification of IoT botnet attacks using deep learning techniques showed good accuracy. These approaches also function well with a variety of feature counts, and in general, more features do not degrade their efficiency, allowing for the use of all data features in a real-world setting. To recognize benign and irregular traffic, this study proposes a proficient anomaly detection technique in view of deep learning for IoT network design. This system actually gains significant complex examples from IoT network streams. suggested to train and test On the recently made available IoT-Botnet 2020 dataset, a deep autoencoder is tested. Several data processing procedures, including feature extraction, data cleaning, and data pre-processing, are carried out to provide the best outcomes. We construct a number of metrics, including Accuracy, Precision, Recall, Confusion Matrix, and FPR, to assess how well our suggested model performs. A comparison between the proposed model and the current RF and MLP approaches is also done as proof. The ML based RF scheme works with an efficiency of 90.0%, and the neural network based scheme MLP shown the accuracy of 83.39%, while the proposed Deep Learning scheme, deep autoencoder has proved its superiority among the other two methods considered in this study.

The data under consideration in this inquiry is N-BaIoT, which presents 115 aspects of the data samples. A botnet is the type of DDoS attack under consideration. The viruses BASHLITE and Mirai are used to cause network anomalies. For the aforementioned considerations, the results reported in this paper are validated. However, there is still need to investigate the performance of deep autoencoders for various types of datasets and malwares.

## REFERENCES

- [1] Harb, H., Mansour, A., Nasser, A., Cruz, E. M., & de la Torre Diez, I. (2020). A sensor-based data analytics for patient monitoring in connected healthcare applications. *IEEE Sensors Journal*, 21(2), 974-984.
- [2] Haider, I., Khan, K. B., Haider, M. A., Saeed, A., & Nisar, K. (2020, November). Automated robotic system for assistance of isolated patients of coronavirus (COVID-19). In *2020 IEEE 23rd International Multipopic Conference (INMIC)* (pp. 1-6). IEEE.
- [3] Hovav, S., & Tsadikovich, D. (2015). A network flow model for inventory management and distribution of influenza vaccines through a healthcare supply chain. *Operations Research for Health Care*, 5, 49-62.
- [4] Sarkar, N. I., Kuang, A. X. M., Nisar, K., & Amphawan, A. (2014). Performance studies of integrated network scenarios in a hospital environment. *International Journal of Information Communication Technologies and Human Development (IJICTHD)*, 6(1), 35-68.

- [5] Mehmood, Y., Ahmad, F., Yaqoob, I., Adnane, A., Imran, M., & Guizani, S. (2017). Internet-of-things-based smart cities: Recent advances and challenges. *IEEE Communications Magazine*, 55(9), 16-24.
- [6] Patel, R., Longini Jr, I. M., & Halloran, M. E. (2005). Finding optimal vaccination strategies for pandemic influenza using genetic algorithms. *Journal of theoretical biology*, 234(2), 201-212.
- [7] Haque, M. R., Tan, S. C., Yusoff, Z., Nisar, K., Lee, C. K., Chowdhry, B. S., ... & Kaspin, R. (2021, January). SDN architecture for UAVs and EVs using satellite: A hypothetical model and new challenges for future. In 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC) (pp. 1-6). IEEE.
- [8] Ahmad, F., Ahmad, Z., Kerrache, C. A., Kurugollu, F., Adnane, A., & Barka, E. (2019, April). Blockchain in internet-of-things: Architecture, applications and research directions. In 2019 International conference on computer and information sciences (ICCIS) (pp. 1-6). IEEE.
- [9] Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies*, 32(1), e4150.
- [10] Ferrag, M. A., Maglaras, L., Moschogiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, 50, 102419.
- [11] Xiaolong, H., Huiqi, Z., Lunchao, Z., Nazir, S., Jun, D., & Khan, A. S. (2021). Soft computing and decision support system for software process improvement: a systematic literature review. *Scientific Programming*, 2021.
- [12] Haque, M. R., Tan, S. C., Yusoff, Z., Nisar, K., Lee, C. K., Kaspin, R., ... & Memon, S. (2021). Automated controller placement for software-defined networks to resist DDoS attacks. *Computers, Materials & Continua*.
- [13] Apostol, I., Preda, M., Nila, C., & Bica, I. (2021). IoT botnet anomaly detection using unsupervised deep learning. *Electronics*, 10(16), 1876.
- [14] Deekshith Shetty, H. C., Varma, M. J., Navi, S., & Ahmed, M. R. Diving Deep into Deep Learning: History, Evolution, Types and Applications.
- [15] Ahmad, Z., Shahid Khan, A., Nisar, K., Haider, I., Hassan, R., Haque, M. R., ... & Rodrigues, J. J. (2021). Anomaly detection using deep neural network for IoT architecture. *Applied Sciences*, 11(15), 7050.
- [16] Darwish, A., Hassanien, A. E., & Das, S. (2020). A survey of swarm and evolutionary computing approaches for deep learning. *Artificial intelligence review*, 53(3), 1767-1812.
- [17] Baig, M. N., Himarish, M. N., Pranaya, Y. C., & Ahmed, M. R. (2018, May). Cognitive architecture based smart homes for smart cities. In 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 461-465). IEEE.
- [18] Dinesh, B., Kavya, B., Sivakumar, D., & Ahmed, M. R. (2019, April). Conforming test of blockchain for 5G enabled IoT. In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1153-1157). IEEE.
- [19] Shen, X., Lin, X., & Zhang, K. (Eds.). (2020). *Encyclopedia of Wireless Networks*. Cham: Springer International Publishing.
- [20] Fahim, M., & Sillitti, A. (2019). Anomaly detection, analysis and prediction techniques in iot environment: A systematic literature review. *IEEE Access*, 7, 81664-81681.
- [21] Cook, A. A., Misirlı, G., & Fan, Z. (2019). Anomaly detection for IoT time-series data: A survey. *IEEE Internet of Things Journal*, 7(7), 6481-6494.
- [22] Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., ... & Zhou, Y. (2017). Understanding the mirai botnet. In 26th USENIX security symposium (USENIX Security 17) (pp. 1093-1110).
- [23] Czyz, J., Kallitsis, M., Gharaibeh, M., Papadopoulos, C., Bailey, M., & Karir, M. (2014, November). Taming the 800 pound gorilla: The rise and decline of NTP DDoS attacks. In Proceedings of the 2014 Conference on Internet Measurement Conference (pp. 435-448).
- [24] Rossow, C. (2014, February). Amplification Hell: Revisiting Network Protocols for DDoS Abuse. In NDSS (pp. 1-15).
- [25] Sanatinia, A., & Noubir, G. (2015, June). Onionbots: Subverting privacy infrastructure for cyber attacks. In 2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (pp. 69-80). IEEE.
- [26] Kwon, J., Lee, J., Lee, H., & Perrig, A. (2016). PsyBoG: A scalable botnet detection method for large-scale DNS traffic. *Computer Networks*, 97, 48-73.
- [27] Shafiq, U., Shahzad, M. K., Anwar, M., Shaheen, Q., Shiraz, M., & Gani, A. (2022). Transfer Learning Auto-Encoder Neural Networks for Anomaly Detection of DDoS Generating IoT Devices. *Security and Communication Networks*, 2022.
- [28] Kang, B. B., Chan-Tin, E., Lee, C. P., Tyra, J., Kang, H. J., Nunnery, C., ... & Kim, Y. (2009, March). Towards complete node enumeration in a peer-to-peer botnet. In Proceedings of the 4th International Symposium on Information, Computer, and Communications Security (pp. 23-34).
- [29] Milojevic, D. S., Kalogeraki, V., Lukose, R., Nagaraja, K., Pruyne, J., Richard, B., ... & Xu, Z. (2002). Peer-to-peer computing.
- [30] Al-Qatf, M., Lasheng, Y., Al-Habib, M., & Al-Sabahi, K. (2018). Deep learning approach combining sparse autoencoder with SVM for network intrusion detection. *Ieee Access*, 6, 52843-52856.

# Machine Learning Techniques to Enhance the Mental Age of Down Syndrome Individuals: A Detailed Review

Irfan M. Leghari<sup>1</sup>, Hamimah Ujir<sup>2</sup>, SA Ali<sup>3</sup>, Irwandi Hipiny<sup>4</sup>

Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak<sup>1, 2, 4</sup>  
Faculty of Artificial Intelligence & Mathematical Sciences, SMIU, Karachi-74000, Pakistan<sup>3</sup>

**Abstract**—Down syndrome individuals are known as intellectually disabled people. Their intellectual ability is classified into four categories known as mild, moderate, severe, and profound. These individuals have significant limitations in learning and adapting skills. Psychologists evaluate mental capability of such individuals using conventional intellectual quotient method instead of using any technology. The research matrix shows most of research has been carried out on analyzing neuroimaging, antenatal screening, and hearing impairment of individuals. But there is still an obvious gap of evaluating mental age using artificial intelligence. We have proposed an artificial neural network model, which supervises how software is used to obtain dataset using Knowledge Base Decision Support System. In a survey (N = 120) individuals examined by psychiatrist, medical expert, and a teacher to assess the presence of Down's syndrome by analyzing their physical and facial appearances, and communication skills. Only (N = 62) individuals declared as Down syndrome. Selected individuals invited to perform mental ability assessment using Interactive Mental Learning Software. The results of mental age of Down syndrome with a raise in IQ from severe to moderate (20% to 35%), moderate to mild (35% to 75%) severity were carried out with the help of assessing the interactive series of software opinion polls based on comparison, logic, and basic mathematical operations using initial IQ (iIQ), and enhanced IQ (eIQ) variables input and output parameters.

**Keywords**—Artificial Intelligence; Artificial Neural Network (ANN); Down Syndrome Individuals (DSI); Interactive Mental Learning Software (IMLS)

## I. INTRODUCTION

The usage of artificial intelligence (AI) in biological fields is increasing, but it's usage in mental disorders is only partial [1-2]. Machine learning (ML) supports the integration of psychological, clinical, and social aspects when approaching the diagnosis of impairments [3]. Artificial intelligence-based applications have promptly been developed for psychiatric diagnosis [4-9]. Furthermore, AI has enormous potential for defining the diagnosis of mental illnesses [10-12]. The purpose of this research is to provide a smart way of learning for Down syndrome individuals (DSI). Of course, this will be helpful for them in practical and professional life, so that they can behave independently with less assistance from their parents or teachers. Until now, research has been carried out on their facial expressions, prediction of their inhibitory capacity, and prediction of mental deficiency using clinical technologies, but still, no work has been done on learning using artificial

intelligence technology, which is very important and needs to be focused on. Artificial intelligence techniques provide the best solution using machine learning and artificial neural networks, this work will be explained in section 4.3.

Born with multiple challenges, Down syndrome individuals are part of every society. They are supposed to be a social and economic burden on families and society. Down syndrome is caused by the presence of an extra copy of chromosome-21. The prevalence level of DSI is approximately 1 in every 800 births [13]. These individuals may have significant cognitive impairments and have an intelligence quotient (IQ) ranging from 30 to 70 percent. In addition, mental abilities that are mostly decreased in these individuals including expressive language, memory, and fine motor skills. Such individuals also have significant limitations in learning and adapting. Adaptive abilities are linked with general mental skills measured with IQ [14]. While the quality of life for DS individuals is improving in both the educational and social domains [15].

The learning process associated with cultural and environmental factors is important for DS individuals due to their social requirements and independency [16-18]. Hence, due to the common difficulties in mental and fine-motor skills, the potential of the individual with Down syndrome as a learner might be perceived as limited [19-21]. They face several different problems in daily life activities while walking, talking, chewing, and learning [20]. The learning ability of DS individuals is classified into four categories: mild, moderate, severe, and profound (Fig. 1). These categories are classified according to their mental age. But in general practice, DS people are grouped as per their physical age. This classification depends on the range of intelligent quotient scores and symptoms [14].

	A normal mind	75 - 100%
a.	Mild	50 - 75%
b.	Moderate	35 - 50%
c.	Severe	20 - 35%
d.	Profound	10 - 20%

Fig. 1. Level of mental impairment of Down Syndrome Individuals.

- 1) Learn in homogeneous educational environments and perform social activities with the least support.
- 2) Individuals need partial support from parents and teachers to carry out tasks. The practitioners use rehabilitation tactics for development ranging from moderate to mild.
- 3) This category requires continued learning and support to carry out an activity.
- 4) Significant deficits in adaptive and functional skills.

The research plan demonstrates the selection criteria using a survey bifurcated into two parts, i.e., general, and technological. After seeking the research gap the research contributions propose the technical way to evaluating unidentified mental age of DS individuals. The research further discusses the implementation of software-based ANN model (Fig. 2).

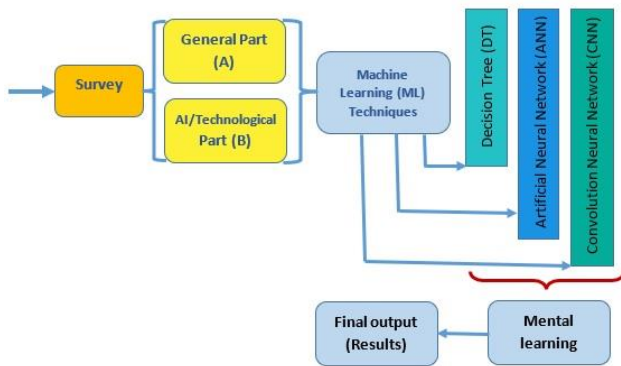


Fig. 2. Proposed research plan.

## II. LITERATURE REVIEW

During the artificial intelligence age of the last decade (2012-2022), an essential number of studies have addressed the use of AI techniques to support people with intellectual disability. The comprehensive literature presents review of the highly illustrative research work of the past ten years i.e., 2012-2022. Furthermore, a research paper matrix is used to identify the research gap in the context of the previous decade (Table I).

Vicari et al., Rosen et al., & Hategan et al., [22-24] proposed a neuroimaging technique using machine learning (ML). Potential bio-supportive artificial intelligence models used to predict mental age are based on brain or neuroimaging data, which serve as neuroimaging data with ML techniques that provide excessive insights for developmental disabilities in Down syndrome children. Machine language models were proposed using imaging recognition features. Research carried out to present knowledge of the neurocognitive, psychopathological, and neurobiological assessment and treatment of patients with Down Syndrome, which suggested rehabilitation as sole effective method for improving cognitive and linguistic abilities.

Amanda Saksida et al., [25] highlighted the problem of cognitive and hearing impairment that happened to most DS children during early childhood. The authors evaluated the effects of hearing impairment on receptive language and

hearing skills and observed main factors of cognitive decline using audiometry testing. In a survey over 41 participants aged between 3 and 10 with DS out of 150 excluding individuals with serious disorders of language, visionary, and cognitive with an IQ of 40, were referred for the audio-logical inspection process. Cognitive skills of 17 individuals of 6 years were measured.

Falin H.E et al., [26] presented a machine learning model to predict Down syndrome in third trimester antenatal screening. The authors used the machine learning (ML) random forest model to predict Down's syndrome. In a survey, around 58,972 pregnant women underwent screening to analyze predictive efficiency. The ML model predicted ratio of 66.7% DS, with a 5% false positive rate in the data set. The model achieved a DS detection rate of 85.2%, with a 5% false positive rate. The study showed that the ML model expands the DS prediction rate with a similar false positive ratio in contrast to the laboratory risk model.

Furthermore, Jojoa-Acosta et al. [27] investigated how does a neuropsychological assessment of intellectual functioning in people with Down syndrome changes over time. The purpose of the research was to predict repressive control capacity using a novel data-driven method. A sample of  $n = 188$ ; 49.47% men; and  $33.6 \pm 8.8$  DS adult individuals having mild-moderate levels of mental retardation was taken into the process. Machine learning Random Forest model used to support vector machine and logistic regression algorithms for the prediction of inhibition capacity. The neuropsychological method was applied for data collection of assessment of memory skills, language skills, executive functions, and praxis was submitted for execution in an algorithm. The outcomes reveal that the finest interpreters for inhibition capacity were verbal memory, constructive praxis, planning, immediate memory, and written verbal comprehension.

In a research led by Children's National Hospital [28] a software device built using machine learning and deep learning technology that detects the presence of the genetic syndrome. The innovation of a software device helps children without any access to specific clinics. The designed software increases access and ML technology to predict the syndrome. The method detects the existence of genetic syndromes using facial photographs. The researchers trained data from 2,800 pediatric DS individuals from different countries.

Similarly, Aida Catic et al. [29] proposed an image processing recognition method to identify affected fetuses early in pregnancy through accurate genetic testing to provide the woman with the preference for the selective continuation of the pregnancy or termination. They intend to replace the traditional process of chromosome photographs with image processing recognition and rule-based classification algorithms. A sample of 2500 pregnant women was collected to determine the figures of maternal levels. All women underwent an ultrasound examination. After the ultrasound examination and maternal blood sample, the blood samples were analyzed using the Prisca software. Artificial Neural Network expert system parameters indicate the tested subject has one of the prenatal syndromes or is healthy.

TABLE I. RESEARCH PAPER MATRIX (IDENTIFYING RESEARCH GAP) MACHINE LEARNING TECHNIQUES TO ENHANCE THE INTELLIGENT QUOTIENT LEVEL OF DOWN SYNDROME INDIVIDUALS

Paper	Author & Year	Topic	Machine learning Techniques/Methodology		
			Decision Tree (DT)	Artificial Neural Network (ANN)	Convolution Neural Network (CNN)
The Influence of Hearing Impairment on Mental Age in Down Syndrome: Preliminary Result	Amanda Saksida et al October 2021	Analyzing whether hearing impairment has a connection with the cognitive problem of Down syndrome individuals.	X	X	X
A machine learning model for the prediction of down syndrome in second-trimester antenatal screening	Falin H.E et al October 2021	Trimester antenatal screening using Machine learning random forest model	X	✓ <sup>1</sup> RF	X
Executive Functioning in Adults with Down Syndrome: Machine-Learning-Based Prediction of Inhibitory Capacity	Jojoa-Acosta et al October 2021	Machine-Learning-Based Prediction of Inhibitory Capacity	X	✓ <sup>1</sup> RF <sup>2</sup> SVM <sup>3</sup> LRA	X
Machine learning tool detects the risk of genetic syndromes in children with diverse backgrounds	Children's National Hospital September 2021	This machine learning technology indicates the presence of a genetic syndrome from a facial photograph	X	✓ <sup>4</sup> DL	X
Application of Neural Networks for classification of Patau, Edwards, Down, Turner and Klinefelter Syndrome based on first-trimester maternal serum screening data, ultrasonographic findings and patient demographics	Aida Catic et al 2018	To identify affected fetuses early in pregnancy through amniocentesis with accurate genetic testing.	X	✓	X
Brain-predicted age in Down syndrome is associated with beta-amyloid deposition and cognitive decline	James H. Cole et al August 2017	Predict brain age using structural neuroimaging data in DS individuals	X	✓ <sup>4</sup> DL	X
Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers	James H.Cole December 2017	Machine learning supervised model for brain age prediction. The predicted brain age was used as a metric to statistically relate to other measured characteristics of the participants	X	✓ <sup>5</sup> NI	X
A pilot study of the use of emerging computer technologies to improve the effectiveness of reading and writing therapies in children with Down syndrome	Vanessa G. Felix et al February 2016	The tool helps to improve reading and writing abilities in Spanish, through mobile computing, multimedia design, and computer speech-recognition techniques named HATLE. During the data collection survey various assessment taken out. Participants were from 6 to 15 years old. <i>IQ scores were not available for any of the participants.</i>	X	X	X
Using Dynamic Bayesian Networks for the Prediction of Mental Deficiency in Children with Down Syndrome	Housseem Turki et al 2014	Proposed a new approach to knowledge extraction from temporal data.	X	✓ <sup>6</sup> DBN	X
Cognition in Down syndrome: a developmental cognitive neuroscience perspective	Jamie O. Edgin et al January 2013	The assessment of several functions of this region seems relatively less impaired than other aspects of cognition. Spatial position and implicit memory are also less affected than an object in location binding or episodic memory.			

<sup>1</sup>Random Forest Model

<sup>2</sup>Support Vector Machine

<sup>3</sup>Logistic regression algorithms

<sup>4</sup>Deep learning<sup>5</sup>Neuroimaging

<sup>6</sup>Dynamic Bayesian Networks

James H. Cole et al. [30] employed a machine learning approach to predict mental age of DS individuals using a structural neuroimaging dataset (N = 46). The chronological age subtracted from predicted age to get a different score of brain-predicted age. The research model analyzed the brain-predicted age calculation at three levels. In the first level, the similarity index of the Gaussian Processes (GP) regression model using a magnetic resonance imaging (MRI) dataset was

collected. In the second level model accuracy was assessed for differentiating brain-predicted age. In the third and fourth levels, testing and brain age were predicted. The authors emphasized the need to examine trajectories of change in DS individuals to get further information about the likelihood of future neurologic decline and negative brain ageing.

Moreover, James H. Cole et al., 2017 [31] analyzed the brain diseases burden of age-associated functional decline. A supervised machine learning model proposed for brain age prediction. Neuroimaging data obtained from MRI scans using machine learning regression model. Cross-validation included 90% of participants and a predicted age of let out of 10%. The predicted mental age was compared with the chronological age of test-set participants. The brain-predicted age difference between brain age and chronological age is assumed to reflect advanced ageing and younger brains. The authors have emphasized that the technical aspects of analyzing brain age are further improved. Neuroimaging brain age measures could be used to evaluate neuroprotective impediments.

To improve the communication ability Vanessa G. Felix et al. [32] developed HATLE application to provide a computer-assisted technique for DS individuals. The data was obtained through a survey of DS participants speaking Spanish aged between 6 and 15 years. IQ scores were not available for any of the participants. The average age of DS individuals was 10.4 years. During the assessment, literacy skills including letter identification, reading, handwriting, and spelling were assessed. A score of all assessments from 0 to 10 was obtained. The training with HATLE was processed group-wise using Android tablets and computers. The outcome of the research reveals that the initial recognition level was set at 0.5, which slowly increased the accuracy rate of further demanding thresholds in steps of 0.1.

Housseem Turki et al. [33] proposed a Dynamic Bayesian Network (DBN) for knowledge extraction from historical data on temporal data to develop a structured learning algorithm for predicting mental retardation in Down syndrome individuals. The experiment took place at the Medical Genetics and Child Psychiatry departments at a hospital in Tunisia. The authors obtained a heterogeneous dataset in collaboration with a team of experts. The purpose of the research is the extraction of knowledge from a great number of datasets that evolve dynamically.

Jamie O. Edgin et al. [34] analyzed a problem with late-developing neural systems in DS individuals and the function of the prefrontal cortex. The assessment of functions was observed relatively less as compared to other aspects of cognition. The results observed were that implicit memory and spatial position are less affected than an object in episodic memory. The authors recommended further study of the fractionated skills patterns in DS individuals, which may benefit developmental change of cognitive functions.

To provide similar learning opportunities for differently abled people Syed Ali [35] proposed a model for adaptation of the Heterogeneous Education System (HES) to the Homogeneous Education System (HES) proposing information technology tools of speech recognition and mathematics. The proposed model suggests that by providing the procedure of conversion and tools, equal opportunities can be provided to different disabilities in the same learning environments. The research has not particularly been done for DS individuals, but the mechanism strongly suggests for all individuals with perform differently. Hence, the research delivers importance to enhance learning and to improve communication difficulties.

In neuroimaging data retrieval, Vicari et al., [22] proposed techniques assessed, including magnetic resonance imaging (MRI), a biotechnology body and brain imaging scanning technology. Rosen et al.; Hategan et al.; Raznahan et al.; & Wintermark et al. [23-24], [36-37] proposed that magnetic imaging is the leading clinical technique to evaluate the level of mental impairment. This technique is used to analyze psychiatric abnormalities that are difficult to detect using computed tomography (CT) For example, AI multimodal learning applications and deep learning methods have been developed for brain imaging [38]. Moreover, convolutional neural networks [39] and deep neural networks [40-42] engaged in neuroimaging to explain the neural relationships of mental disorders [40] [43-46].

Heinsfeld et al., [47] proposed that electroencephalography (EEG) signals are important to understanding how the human mind works and evaluating mental impairment. In contrast to MRI and CT, electroencephalography has greater resolution [48] analysed by Grotegerd et al. In addition, EEG data graphs were evaluated using artificial intelligence models presented by Hannesdóttir et al.; Avram et al., Thibodeau et al., & Hosseinifard [49-52].

### III. RESEARCH CONTRIBUTIONS

#### A. Mental Age Evaluation

This research illustrated the valuable studies that tried to solve problems in evaluating and diagnosing mostly researched cognitive impairment, i.e., Down syndrome. Wherein, artificial intelligence neural network model-based techniques and software approaches are implemented to bring down syndrome analysis and seek ratio of their mental approach and to further strengthen them with software. The proposed model (Fig. 3) is divided into two portions (*A and B*). Firstly, the model evaluates the identification of Down syndrome. Here, the model reveals three major components of cognitive psychology known as cognitive neuroscience, human psychology, and information processing through computers. As per the neuroscience perspective, thinking abilities depend on working memory. The area of cognitive psychology considers the study of mental functions in which people require knowledge to understand their experiences. The model emphasizes both artificial intelligence and biological methods. The investigation is applied to individuals diagnosed with Down syndrome.

A survey comprised over 120 individuals with DS of different age groups ( $\geq 8$  &  $\leq 30$ ) was included to identify different cognitive traits (Fig. 5). The survey was based on interviews conducted by a team consisting of a psychiatrist, special education instructor and parents accompanying the DS individual. The team of psychiatrists, based on observations and professional knowledge, identified intellectual disability using facial expressions and psychological traits of DS individuals. Distinct facial features include distinctive slanting eyes, a small chin, abnormal outer ears, a flat nasal bridge, and a flattened nose. Psychological traits include talking, paying attention, and social rules. A team of special education teachers used simple mathematical problems to evaluate numerical skills, reasoning, and decision-making skills. The research contributions are further based on an artificial neural network model to evaluate the intelligent quotient of Down syndrome

individuals. The intelligent algorithm reveals the criteria of the artificial neural network model. The software access repeatedly until their mental functioning improves from severe to moderate and from moderate to mild levels using variables initial IQ (iIQ), and enhanced IQ (eIQ) denoted as input and output parameters based on the practical, creative, and analytical testing. The method constructs membership functions building set of rules into the knowledge base and evaluates rules in the Inference Engine (Table II).

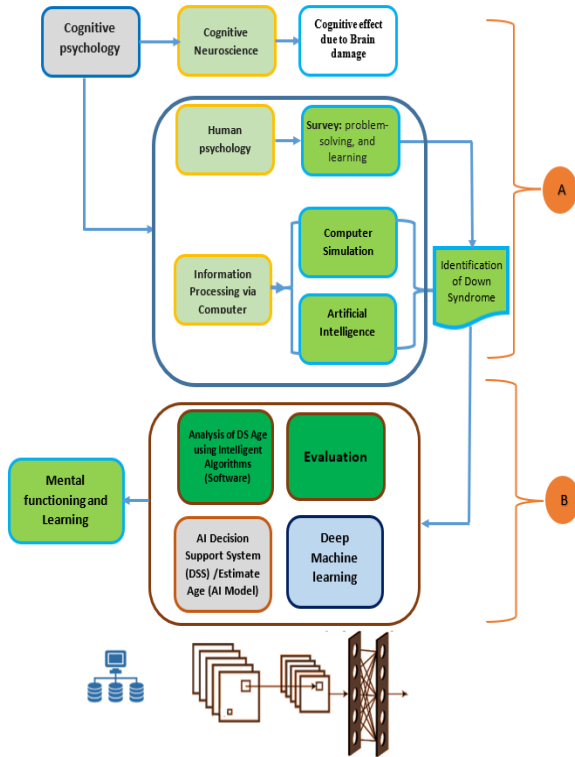


Fig. 3. A proposed model of mixed approach.

The artificial neural network model presents the mechanism of the intelligent algorithm for repeating the usability of the interactive software, denoted as the middle layer of the ANN model (Fig. 4). The neural network model is supposed to judge the cognitive traits of individuals to analyze the IQ. The middle layer plays a part in the manifest of the interactive software application. The overall process is based on three layers. The input layer holds the initial data of the learning process of DS individuals (8–30 years). The intermediate hidden layer represents long-term memory, sensing, decision-making, perception, supervisory skills, thinking, logic, and learning complexities of the DS individuals. The hidden layer performs a nonlinear transformation of the inputs entering the network. The computations from the hidden layer are transformed into the output layer to reveal data to the outside world in the form of the computed mental level and capability.

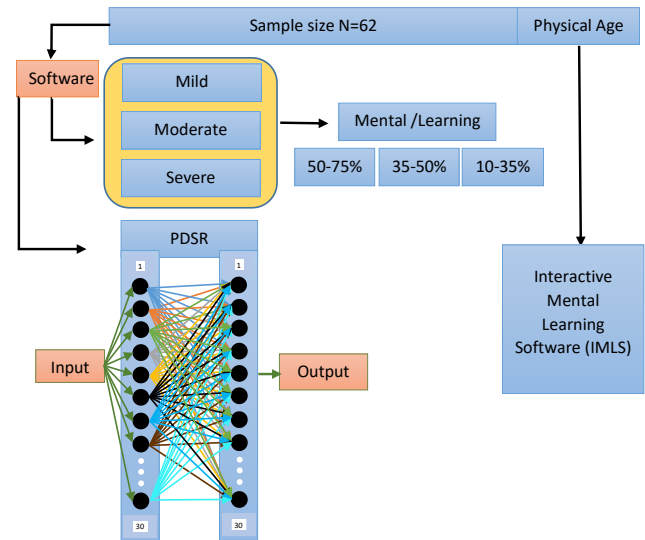


Fig. 4. Artificial neural network model.

TABLE II. INTELLIGENT ALGORITHMS OF INTERACTIVE MENTAL LEARNING SOFTWARE

<p><i>Step 1 – Define variables and parameters</i>  <i>Variables or parameters are inputs and their outputs like parameter for IQ are,</i>  <i>IQ (iq) = {practical, creative, analytical}</i>  <i>Every member of this set is covering some level of intelligence IQ values.</i>  <b>Create a matrix of based on their syndrome and mental impairment. Build a set of rules into the knowledge base in the form of IF-THEN-ELSE structures.</b>                  If {level of Down Syndrome is <b>Mild</b>} then (IQ level <b>50-75</b>) Else                  If {level of Down Syndrome is <b>Moderate</b>} then (IQ level <b>35-50</b>) Else                  If {level of Down Syndrome is <b>Severe</b>} then (IQ level <b>20-35</b>) Else                  If {level of Down Syndrome is <b>Profound</b>} then (IQ level <b>10-20</b>)</p> <p><i>Step 2 – Construct membership functions for Step 1</i>  <i>Step3 – Maintained knowledge base rules</i>  <i>Step 4 – Evaluate rules in the rule base (Inference Engine)</i>                  End</p>
---

**B. Supervised Learning Algorithm**

We have implemented ML algorithms to evaluate the projected mental age of the DS individuals. Individuals with Down syndrome's brain/mental age efficiency improves with practice on a software based on supervised learning algorithm. The end of the exponential research (N = 120), an efficient framework is proposed to identify and improve the mental age of Down syndrome children and young adults (Table III). For the analysis of training set sizes, bootstrapping techniques are used to estimate the reliability of the ML algorithm for different training set sizes. In the ANN technique, the resampling method is used to resample the original training set with a replacement to get a new training set of the chosen sample size. The advantage of this technique is that it allows us to judge the toughness of performance around training dataset sizes and to recognize the smallest training set sizes essential for checking the performance above the expected level.

**C. Multicriteria Decision Support System (DSS)**

To improve the learning abilities of DS individuals, the Knowledge Base Decision Support System (DBSS) is used in cascading fashion. The decision support system is based on four multiple criteria for DS individuals for learning and problem solving. The decision support system checks all the four criteria of Down syndrome in cascade and fixes the criteria based on the minimum criterion. The algorithm is based on decision-based criteria over an alleged cascading in descending order of the learning model on minimum criteria.

- Call at the Dataset.
- A cascading effect scenario to detect.
- Decision support algorithm is used for deciding the set accordingly to minimum criteria.
- Intelligent Quotient (IQ).

**D. Interactive Mental Learning Software**

The interactive graphical environment is based on 30 public opinion polls (Table IV). Software provides an interactive and simple platform to use opinion pool with the help of teacher or parents. The series of opinion polls contain questions based on comparison, logic, and basic mathematical operations. Individuals are supposed to search for the best option to increase their mental score at the end of the pool (Fig. 6).

The software is accessible through laptop, desktop, or smartphone. Input is selected using a mouse, keypad or by touchpad. Different series of questions appears showing three options to choose best one. Colorful shapes of birds, fruit, and vegetables, colors, vehicles enhance interest of individuals and reduce frustration. The comparison covers the questions of the basic shapes and figures, which helps in developing the logic of the DS individuals. Making comparisons between numbers and alphabets helps DS individuals develop their decision-making abilities. Basic mathematical operations cover only addition (+) operation. The mathematical console is comprised of the addition of birds, animals, shapes, and numbers. The assessment process is divided into three rounds (Round-I, II &

III). Round-I process (N = 20) individuals (Table V). Round-II processed (N = 20) (Table VI) and Round-II processed (N = 22) individuals (Table VII).

TABLE III. ARCHITECTURE FOR MENTAL AGE (MA) OF DS INDIVIDUALS

Age assessment Module		System	Inference Engine (Decision Support System)	System Module
Mild	50-75	It stores IF-THEN rules provided by experts.	It simulates the human reasoning process by making system inference on the inputs and IF-THEN rules.	It transforms the set obtained by the inference engine into a crisp value.
Moderate	35-50			
Severe	20-35			
Profound	10-20			

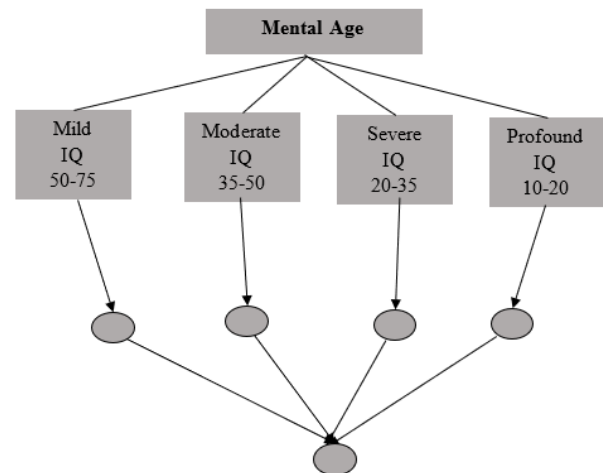


Fig. 5. Decision tree of Down Syndrome individuals.

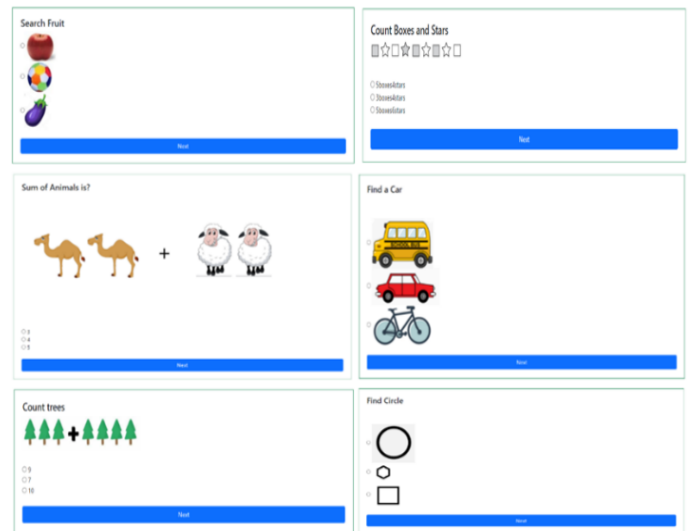


Fig. 6. Interactive Mental Learning Software (IMLS).



TABLE IV. INTERACTIVE MENTAL LEARNING (IML) SOFTWARE CRITERIA

Q #	Question (Pictorial/Text)	Options given in pictorial / text form
Q1	Search Fruit (images)	(a) Fruit (b) Ball (c) Vegetable
Q2	Click Red Color (Colors)	(a) Aqua (b) Red (c) Yellow
Q3	Find a number (5)	(a) 5 (b) M (c) A
Q4	Find a Car (images)	(a) Bus (b) Car (c) Bicycle
Q5	Search greater number	(a) 10 (b) 5 (c) 0
Q6	Search 3 Birds (images)	(a) 3 birds (b) 2 birds (c) 1 bird
Q7	Find a Sheep (images)	(a) Camel (b) Sheep (c) Goat
Q8	Count Donuts (images: 8)	(a) 6 (b) 7 (c) 8
Q9	Sum of Animals is (images: 2 Camel and 2 sheep)	(a) 3 (b) 4 (c) 5
Q10	M for:	(a) Jeep (b) Car (c) Mobile
Q11	A for:	(a) Apple (b) Banana (c) Cat
Q12	1 + 1	(a) 3 (b) 2 (c) 4
Q13	Which Bird is Flying? (images)	(a) 1 Nonflying bird (b) 2 NFB (c) 1 Flying Bird
Q14	We go to school by. (images)	(a) Car (b) Bus (c) Bicycle
Q15	Rabbit lives in? (images)	(a) Human House (b) Tree (c) Burrows
Q16	Rabbit eats? (images)	(a) Donut (b) Muffin (c) Carrot
Q17	Cat run after? (images)	(a) Bird (b) Dog (c) Rat
Q18	Goat gives? (images)	(a) Eggs (b) Milk (c) Fish
Q19	Aisha is a female?	(a) Yes (b) No
Q20	We fly in. (images)	(a) Car (b) Ship (c) Aeroplane
Q21	Count small circles?	(a) 8 (b) 9 (c) 7
Q22	Count Stars? (x + xx)	(a) 4 (b) 2 (c) 3
Q23	Count Circles? (images)	(a) 2 (b) 3 (c) 4
Q24	Count Boxes & Stars	(a) 5boxes4stars (b) 3boxes4stars
Q25	We talk on. (images)	(a) Mobile (b) Watch (c) Laptop
Q26	Count trees? (image)	(a) 9 (b) 7 (c) 10
Q27	Count clouds? (image)	(a) 10 (b) 11 (c) 12
Q28	Find a Circle? (images)	(a) Circle (b) Hexagon (c) Square
Q29	Find a Square? (images)	(a) Square (b) Hexagon (c) Circle
Q30	Which circle is big? (images)	(a) Small circle (b) Big circle

TABLE V. SOFTWARE DRIVEN AGE ROUND-I SAMPLE SIZE N=20 MILD (50-75%) MODERATE (35-50%) SEVERE (20-35%) MODERATE (10-20%)

S.No	Participants	Physical Age (Years)	Processed Mental level	Improved traits
1.	Participant-1 (Male)	12	45%	Perception (40%), attention (45%)
2.	Participant-2 (Female)	13	50%	Sensing (50%), Reasoning (45%)
3.	Participant-3 (Male)	15	25%	Sensing (25%), memory (20%)
4.	Participant-4 (Female)	15	45%	Responsive (40%), attention (45%)
5.	Participant-5 (Male)	14	50%	Decision making (50%), Reasoning (50%)
6.	Participant-6 (Female)	8	45%	Perception (40%), attention (45%)
7.	Participant-7 (Male)	17	65%	Sensing (65%), memory (65%)
8.	Participant-8 (Female)	8	60%	Sensing (60%), Logic (55%)
9.	Participant-9 (Female)	13	65%	Perception (65%), memory (55%)
10.	Participant-10 (Male)	10	55%	Logic (50%), attention (55%)
11.	Participant-11 (Male)	12.6	45%	Responsive (45%), memory (40%)
12.	Participant-12 (Male)	8	30%	Responsive (30%), attention (35%)
13.	Participant-13 (Male)	9	35%	Attention (35%), DM (35%)
14.	Participant-14 (Male)	18	70%	Reasoning (65%), Memory (70%)
15.	Participant-15 (Male)	15	55%	Social (50%), attention (55%)
16.	Participant-16 (Female)	18	55%	Logic (50%), attention (55%)

17.	Participant-17 (Male)	9	45%	Sensing (45%), FM Skills (45%)
18.	Participant-18 (Male)	16	65%	Reasoning (60%), attention (65%)
19.	Participant-19 (Female)	9	55%	Sensing (50%), Responsive (55%)
20.	Participant-20 (Female)	11	35%	Logic (30%), Memory (35%)

TABLE VI. SOFTWARE DRIVEN AGE ROUND-II SAMPLE SIZE N=20 MILD (50-75%) MODERATE (35-50%) SEVERE (20-35%) MODERATE (10-20%)

S.No	Participants	Physical Age (Years)	Processed Mental level	Response toward activities
1.	Participant-21 (Female)	10	30%	Attention - memory (30%)
2.	Participant-22 (Female)	12	55%	Sensing (55%), DM (50%)
3.	Participant-23 (Female)	14	45%	Perception (45%), memory (40%)
4.	Participant-24 (Male)	11	55%	Sensing - attention (55%)
5.	Participant-25 (Female)	15	55%	Logic (50%), DM (55%)
6.	Participant-26 (Female)	13	45%	VM (45%), attention (40%)
7.	Participant-27 (Female)	14	35%	Sensing – Logic (30%)
8.	Participant-28 (Male)	10	25%	Reasoning (30%), Memory (25%)
9.	Participant-29 (Male)	18	50%	Reasoning - attention (55%)
10.	Participant-30 (Female)	16	45%	Logic - attention (45%)
11.	Participant-31 (Female)	8	35%	Sensing - FM Skills (35%)
12.	Participant-32 (Male)	13	55%	Reasoning - attention (55%)
13.	Participant-33 (Female)	12	40%	Sensing (40%), Responsive (45%)
14.	Participant-34 (Male)	10	55%	Logic - Responsive (55%)
15.	Participant-35 (Male)	16	35%	Sensing - memory (30%)
16.	Participant-36 (Female)	8	40%	Sensing (40%), memory (35%)
17.	Participant-37 (Female)	14	65%	Reasoning - memory (65%)
18.	Participant-38 (Female)	9	40%	Sensing (40%), memory (45%)
19.	Participant-39 (Male)	10	55%	Decision making-responsive (50%)
20.	Participant-40 (Female)	15	45%	Sensing (45%), memory (50%)

TABLE VII. SOFTWARE DRIVEN AGE ROUND-III SAMPLE SIZE N=22 MILD (50-75%) MODERATE (35-50%) SEVERE (20-35%) MODERATE (10-20%)

S.No	Participants	Physical Age (Years)	Processed Mental level	Improvements in cognitive traits
1.	Participant-41 (Male)	15	45%	Social – Decision making (40%)
2.	Participant-42 (Male)	25	60%	Reasoning (60%), DM (55%)
3.	Participant-43 (Male)	30	65%	Logic (50%), Memory (65%)
4.	Participant-44 (Female)	20	60%	Sensing – Decision making (55%)
5.	Participant-45 (Female)	22	40%	Attentive (40%), Memory (45%)
6.	Participant-46 (Male)	15	35%	Sensing – Logic (35%)
7.	Participant-47 (Male)	23	55%	Sensing (55%), Memory (60%)
8.	Participant-48 (Male)	27	70%	Social (70%), Sensing (65%)
9.	Participant-49 (Male)	25	45%	Attention (40%), Sensing (45%)
10.	Participant-50 (Female)	30	50%	Memory (50%), Attention (45%)
11.	Participant-51 (Male)	22	60%	Social (70%), Sensing (65%)
12.	Participant-52 (Male)	25	55%	DM (55%), FM Skills (50%)
13.	Participant-53 (Male)	28	45%	Social (40%), Memory (45%)
14.	Participant-54 (Male)	27	60%	Reasoning – Attention (60%)
15.	Participant-55 (Female)	22	45%	DM (45%), Sensing (40%)
16.	Participant-56 (Female)	20	50%	Memory (50%), DM (45%)
17.	Participant-57 (Male)	30	45%	Attention (40%), Sensing (45%)
18.	Participant-58 (Female)	19	50%	Social – Decision making (55%)
19.	Participant-59 (Male)	17	60%	Sensing (60%), Memory (65%)
20.	Participant-60 (Female)	17	45%	Responsive (45%), Sensing (40%)
21.	Participant-61 (Female)	30	50%	Logic (50%), Sensing (45%)
22.	Participant-62 (Female)	25	50%	Logic (50%), Memory (45%)

#### IV. DISCUSSION AND FUTURE WORK

The research carried out to investigate the mental age of DS individuals. The literature matrix identified the research gap of assessing actual mental age using AI-ANN model. The technology claims to enhance mental age of DS individuals having least IQ level. The interactive and simplest platform of software increase usability interest and reduce frustration. The different question of logical, mathematical, and analytical reasoning boosts the thinking ability, perception, reasoning, logic, and memory of the individuals. Such traits result change in IQ from severe to moderate (IQ>20% to 35%), moderate to mild (IQ>35% to 75%). Research outcomes also show the comparison and authenticity between software-based IQ assessment and traditional methods. The variation in the mental age is identified with yellow line of the graphical illustration (Fig.7, 8 and 9).

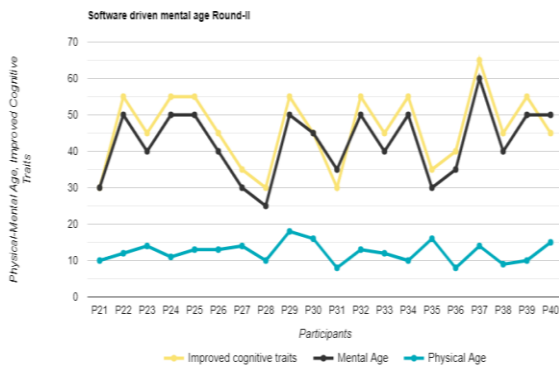


Fig. 7. Graph - software driven mental age round-I.

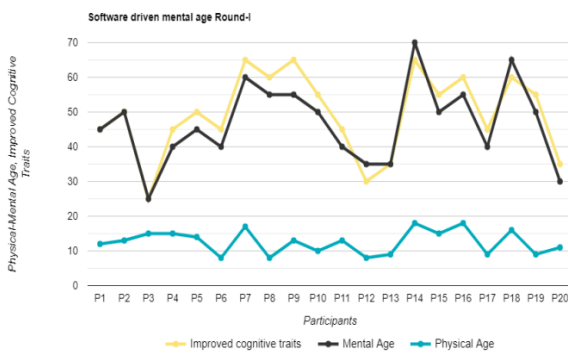


Fig. 8. Graph - software driven mental age round-II.

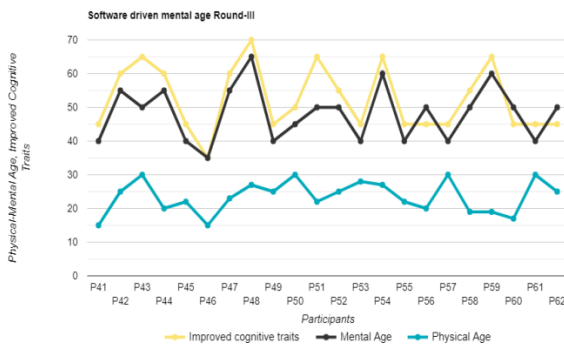


Fig. 9. Graph - software driven mental age round-III.

In future work, more feature added system for enhancement of severe level of mental retardation may be suggested for academic purpose to help different intellectual disabilities in special education and job oriented technical training to make them independent.

#### ACKNOWLEDGMENT

The authors fully acknowledged Universiti Malaysia Sarawak for all the support which makes this important research viable and effective. The acknowledgement is also extended to all participants who were voluntary involved in the dataset collection.

#### REFERENCES

- Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J., "Artificial intelligence in radiology", in Nature Reviews Cancer, 2018, 18(8), pp. 500-510.
- Ray, A., Bhardwaj, A., Malik, Y. K., Singh, S., & Gupta, R., "Artificial intelligence and Psychiatry: An overview", in Asian Journal of Psychiatry, 2022, 103021.
- Graham, S. A., Lee, E. E., Jeste, D. V., Van Patten, R., Twamley, E. W., Nebeker, C., ... & Depp, C. A., "Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: A conceptual review", in Psychiatry research, 2020, 284, 112732.
- Liu, G. D., Li, Y. C., Zhang, W., & Zhang, L., "A brief review of artificial intelligence applications and algorithms for psychiatric disorders", in Engineering, 2020, 6(4), pp. 462-467.
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J., "A guide to deep learning in healthcare", in Nature medicine, 2019, 25(1), pp. 24-29.
- Dwyer, D. B., Falkai, P., & Koutsouleris, N., "Machine learning approaches for clinical psychology and psychiatry", in Annual review of clinical psychology, 2018, 14, pp. 91-118.
- Sarma, G. P., Hay, N. J., & Safran, A., "AI safety and reproducibility: establishing robust foundations for the neuropsychology of human values", in International Conference on Computer Safety, Reliability, and Security, Springer, Cham, 2018, pp. 507-512..
- Stead, W. W., "Clinical implications and challenges of artificial intelligence and deep learning", in Jama, 2018, 320(11), pp. 1107-1108.
- Gao, H., Yin, Z., Cao, Z., & Zhang, L., "Developing an agent-based drug model to investigate the synergistic effects of drug combinations", in Molecules, 2017, 22(12), 2209.
- Xia, Y., Yang, C., Hu, N., Yang, Z., He, X., Li, T., & Zhang, L., "Exploring the key genes and signaling transduction pathways related to the survival time of glioblastoma multiforme patients by a novel survival analysis model", in BMC genomics, 2017, 18(1), pp.1-11.
- Zhang, L., Xiao, M., Zhou, J., & Yu, J., "Lineage-associated underrepresented permutations (LAUPs) of mammalian genomic sequences based on a Jellyfish-based LAUPs analysis application (JBLA)", in Bioinformatics, 2018, 34(21), pp. 3624-3630.
- Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H. C., & Jeste, D. V., "Artificial intelligence for mental health and mental illnesses: an overview", in Current psychiatry reports, 2019, 21(11), pp. 1-18.
- Jamshidnezhad, A., Hosseini, S. M., Mohammadi-Asl, J., & Mahmudi, M., "An intelligent prenatal screening system for the prediction of Trisomy-21", in Informatics in Medicine Unlocked, 2021, 24, 100625.
- Hamburg, S., Lowe, B., Startin, C. M., Padilla, C., Coppus, A., Silverman, W., ... & Strydom, A., "Assessing general cognitive and adaptive abilities in adults with Down syndrome: a systematic review", in Journal of neurodevelopmental disorders, 2019, 11(1), pp. 1-16.
- Baburamani, A. A., Patkee, P. A., Arichi, T., & Rutherford, M. A. "New approaches to studying early brain development in Down syndrome", in Developmental medicine & child neurology, 2019, 61(8), pp. 867-879.

- [16] Barbosa, R. T. D. A., de Oliveira, A. S. B., de Lima Antão, J. Y. F., Crocetta, T. B., Guarnieri, R., Antunes, T. P. C., ... & de Abreu, L. C. "Augmentative and alternative communication in children with Down's syndrome: a systematic review", in *BMC pediatrics*, 2018, 18(1), pp. 1-16.
- [17] Luna-García, H., Mendoza-González, A., Mendoza-Gonzalez, R., Gamboa-Rosales, H., Galván-Tejada, J. I., Celaya-Padilla, J. M., ... & Lopez-Veyna, J. "Analyzing typical mobile gestures in mHealth applications for users with Down syndrome", in *Mobile Information Systems*, 2018.
- [18] Kamoun, P. "Mental retardation in Down syndrome: a hydrogen sulfide hypothesis", in *medical hypotheses*, 2001, 57(3), pp. 389-392.
- [19] Marques, L. S., Alcântara, C. E. P., Pereira, L. J., & Ramos-Jorge, M. L. "Down syndrome: a risk factor for malocclusion severity?", in *Brazilian oral research*, 2015, 29(1), pp. 1-7.
- [20] Jan Blacher, AAMR, "Families and Mental Retardation", in *A collection of Notable AAMR Journal Articles across the 20th century*, 2002.
- [21] Smith, P., & Smith, L. "Artificial intelligence and disability: too much promise, yet too little substance?", in *AI and Ethics*, 2021, 1(1), pp. 81-86.
- [22] Vicari, S., Pontillo, M., & Armando, M. "Neurodevelopmental and psychiatric issues in Down's syndrome: assessment and intervention", in *Psychiatric genetics*, 2013, 23(3), pp. 95-107.
- [23] Rosen BR, Huang SY, Stuffelbeam SM. "Pushing the limits of human neuroimaging", in *JAMA*, 2015, 314(10): pp. 993-4.
- [24] Hategan A, Bourgeois JA, Cheng T, Young J. "Neuropsychology and neuroimaging in clinical geriatric psychiatry", in *Geriatric psychiatry study guide*. Cham: Springer, 2018, pp. 23-38.
- [25] Saksida, A., Brotto, D., Pizzamiglio, G., Bianco, E., Bressan, S., Feresin, A., ... & Orzan, E. "The Influence of Hearing Impairment on Mental Age in Down Syndrome: Preliminary Results", in *Frontiers in Pediatrics*, 2021, 9.
- [26] He, F., Lin, B., Mou, K., Jin, L., & Liu, J. "A machine learning model for the prediction of down syndrome in second trimester antenatal screening", in *Clinica Chimica Acta*, 2021, 521, pp. 206-211.
- [27] Jojoa-Acosta, M. F., Signo-Miguel, S., Garcia-Zapirain, M. B., Gimeno-Santos, M., Méndez-Zorrilla, A., Vaidya, C. J., ... & Bruna-Rabassa, O. "Executive Functioning in Adults with Down Syndrome: Machine-Learning-Based Prediction of Inhibitory Capacity", in *International Journal of Environmental Research and Public Health*, 2021, 18(20), 10785.
- [28] Children's National Hospital, "Machine learning tool detects the risk of genetic syndromes in children with diverse backgrounds: The technology can potentially reduce health inequality in under-resourced areas worldwide", in *ScienceDaily*. Retrieved February 17, 2022, from [www.sciencedaily.com/releases/2021/09/210901191416.htm](http://www.sciencedaily.com/releases/2021/09/210901191416.htm).
- [29] Catic, A., Gurbeta, L., Kurtovic-Kozaric, A., Mehmedbasic, S., & Badnjevic, A. "Application of Neural Networks for classification of Patau, Edwards, Down, Turner and Klinefelter Syndrome based on first trimester maternal serum screening data, ultrasonographic findings and patient demographics", in *BMC medical genomics*, 2018, 11(1), pp. 1-12.
- [30] Cole, J. H., Annus, T., Wilson, L. R., Remtulla, R., Hong, Y. T., Fryer, T. D., ... & Holland, A. J., "Brain-predicted age in Down syndrome is associated with beta amyloid deposition and cognitive decline", in *Neurobiology of aging*, 2017, 56, pp. 41-49.
- [31] Cole, J. H., & Franke, K. "Predicting age using neuroimaging: innovative brain ageing biomarkers", in *Trends in neurosciences*, 2017, 40(12), pp. 681-690.
- [32] Felix, V. G., Mena, L. J., Ostos, R., & Maestre, G. E. "A pilot study of the use of emerging computer technologies to improve the effectiveness of reading and writing therapies in children with Down syndrome", in *British Journal of Educational Technology*, 2017, 48(2), pp. 611-624.
- [33] Turki, H., Ayed, M. B., & Alimi, A. M. "Using dynamic Bayesian networks for the prediction of mental deficiency in children with down syndrome", in *2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPar)*, 2014, pp. 163-167, IEEE.
- [34] Edgin, J. O. "Cognition in Down syndrome: a developmental cognitive neuroscience perspective", in *Wiley Interdisciplinary Reviews: Cognitive Science*, 2013, 4(3), pp. 307-317.
- [35] Ali, S. A., & Burney, S. A. "Conversion of heterogeneous education system (HOES) into homogeneous education system (HOES) for ease of disabled persons using information technology", in *2010 International Conference on Computer Design and Applications*, 2010, (Vol. 2, pp. V2-298). IEEE.
- [36] Park MTM, Raznahan A, Shaw P, Gogtay N, Lerch JP, Chakravarty MM. "Neuroanatomical phenotypes in mental illness: identifying convergent and divergent cortical phenotypes across autism", in *ADHD and schizophrenia. JPsychiatry Neurosci*, 2018, 43(3), pp. 201-12.
- [37] Wintermark M, Colen R, Whitlow CT, Zaharchuk G. "The vast potential and bright future of neuroimaging", in *Br J Radiol*, 2018, 91(1087), 20170505.
- [38] Webb S. "Deep learning for biology", in *Nature*, 2018, 554(7693), pp. 555-7.
- [39] Wolfers T, Buitelaar JK, Beckmann CF, Franke B, Marquand AF. "From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics", in *Neurosci Biobehav Rev*, 2015, 57, pp. 328-49.
- [40] Arbabshirani MR, Plis S, Sui J, Calhoun VD. "Single subject prediction of brain disorders in neuroimaging: promises and pitfalls", in *Neuroimage*, 2017, 145(PtB), pp. 137-65.
- [41] Vieira S, Pinaya WHL, Mechelli A. "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications", in *Neurosci Biobehav Rev*, 2017, 74(Pt A), pp. 58-75.
- [42] Bengio Y. "Learning deep architectures for AI. Hanover", in *Now Publishers Inc.*, 2009.
- [43] LeCun Y, Bengio Y, Hinton G. "Deep learning", in *Nature*, 2015, 521(7553), pp. 436-44.
- [44] Plis SM, Hjelm DR, Salakhutdinov R, Allen EA, Bockholt HJ, Long JD, "Deep learning for neuroimaging: a validation study", in *Front Neuroscience*, 2014, 8:229.
- [45] Payan A, Montana G. "Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks", in *arXiv*, 2015, 1502.02506.
- [46] Hosseini-Asl E, Ghazal M, Mahmoud A, Aslantas A, Shalaby AM, Casanova MF, et al. "Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network", in *Front Biosci*, 2018, 23, pp. 584-96.
- [47] Heinsfeld AS, Franco AR, Craddock RC, Buchweitz A, Meneguzzi F. "Identification of autism spectrum disorder using deep learning and the ABIDE dataset", in *Neuroimage Clin*, 2018, Vol. 17, pp. 16-23.
- [48] Grotegerd D, Suslow T, Bauer J, Ohrmann P, Arolt V, Stuhrmann A, et al. "Discriminating unipolar and bipolar depression by means of fMRI and pattern classification: a pilot study", in *Eur Arch Psychiatry Clin Neuroscience*, 2013, 263 (2), pp. 119-31.
- [49] Hannesdóttir DK, Doxie J, Bell MA, Ollendick TH, Wolfe CD. "A longitudinal study of emotion regulation and anxiety in middle childhood: associations with frontal EEG asymmetry in early childhood", in *Dev Psychobiol*, 2010, 52 (2): pp. 197-204.
- [50] Avram J, Baltes FR, Miclea M, Miu AC. "Frontal EEG activation asymmetry reflects cognitive biases in anxiety: evidence from an emotional face Stroop task", in *Appl Psychophysiol Biofeedback*, 2010, 35(4), pp. 285-92.
- [51] Thibodeau R, Jorgensen RS, Kim S. "Depression, anxiety, and resting frontal EEG asymmetry: a meta-analytic review", in *J Abnorm Psychol*, 2006, 115 (4), pp. 715-29.
- [52] Hosseinifard B, Moradi MH, Rostami R. "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal", in *Computer Methods Programs Biomed*, 2013, 109(3), pp. 339-45.

# AMIM: An Adaptive Weighted Multimodal Integration Model for Alzheimer's Disease Classification

Dewen Ding<sup>1</sup>, Xianhua Zeng<sup>2</sup>, Xinyu Wang<sup>3</sup> and Jian Zhang<sup>4</sup>  
Chongqing University of Posts and Telecommunications  
Chongqing Key Laboratory of Image Cognition, China, Chongqing

**Abstract**—Alzheimer's disease (AD) is an irreversible neurological disorder, so early medical diagnosis is extremely important. Magnetic resonance imaging (MRI) is one of the main medical imaging methods used clinically to detect and diagnose AD. However, most existing computer-aided diagnostic methods only use MRI slices for model architecture design. They ignore informational differences between all slices. In addition, physicians often use multimodal data, such as medical images and clinical information, to diagnose patients. The approach helps physicians to make more accurate judgments. Therefore, we propose an adaptive weighted multimodal integration model (AMIM) for AD classification. The model uses global information images, maximum information slices and clinical information as data inputs for the first time. It adopts adaptive weights integration method for classification. Experimental results show that our model achieves an accuracy of 99.00% for AD versus normal controls (NC), and 82.86% for mild cognitive impairment (MCI) versus NC. The proposed model achieves best classification performance in terms of accuracy, compared with most state-of-the-art methods.

**Keywords**—MRI; global information images; maximum information slices; adaptive weights; integration method

## I. INTRODUCTION

Alzheimer's disease, a chronic neurodegenerative disease causing the death of nerve cells and tissue loss throughout the brain, usually starts slowly and worsens over time. AD is expected to affect 1 out of 85 people in the world by the year 2050 [1]. The progression of AD gradually leads to memory degradation and impairment of cognitive function, eventually leading to irreversible neuronal damage [2]. Although no treatment has been proven to be effective in preventing the progression of AD [3], the early diagnosis of AD remains important for subsequent treatment to delay the onset of cognitive symptoms [4].

Since 2013, deep learning has begun to gain considerable attention in AD detection research, with the number of published papers in this area increasing drastically since 2017 [5]. Early unsupervised methods used autoencoders or restricted Boltzmann machine methods to extract features that were then used for the classification of Alzheimer's disease [6]–[8]. Supervised learning applied to the diagnosis of Alzheimer's disease has been particularly well studied compared to unsupervised methods. Convolutional neural networks (CNNs) are the most successful deep models for image analysis, aiming to make better use of spatial information by taking 2D/3D images as input and extracting features by stacking several

convolutional layers; the result is a hierarchy of progressively abstracted features [9], [10]. Most studies on Alzheimer's disease have been mainly architected by 2D CNNs or 3D CNNs as depth models. A large number of studies have performed feature extraction of MRI slices by 2D CNNs for ADNI classification [11]–[16]. Since MRI provides 3D images, how to select MRI slices is a question worth thinking about. Meanwhile, 2D slices cannot contain all the information of 3D MRI, so it is missing the global information. 3D CNN is widely used for diagnosis of 3D MRI, which does not require slice selection and also contains global information. However, AD detection must take the whole image or some ROI as input [17], [18]. This results in a steep increase in the number of parameters, which can create problems such as large amount of computation, time-consuming, and overlapping data. The joint 2D-3D CNN [19]–[21] first performs 3D feature extraction by multiple 3D data inputs, and then obtains the final classification result by 2D CNN. Likewise, there are also problems such as large amount of computational cost and time consuming. In addition, in the early detection of Alzheimer's disease, the degree of brain atrophy is less variable, and assessment by a single modality of MRI alone may have a certain bias; a combined assessment with multiple modalities will yield a relatively more accurate diagnosis.

Based on slice-level classification, there is a lack of three-dimensional spatial information and the subjective uncertainty of slice selection. We proposed to effectively superimpose all slices to generate a dynamic 2D picture containing multiple slice information changes, i.e. a global information image. At the same time, the slice with the largest amount of information is also selected through the method of image entropy, and the clinical information is used as the input of the multi-dimensional feature auxiliary integrated approach. From our experiment, the proposed AMIM model has improved the performance significantly. The main contributions of our study are summarized in the following three folds.

- We propose an adaptive weighted multimodal ensemble model. The model uses an adaptive weighted method to optimize the different branches weights. It can effectively reduce the large amount of computational cost and time, compared with the grid search method.
- For the first time, we propose a new MRI image preprocessing method, which uses dynamic images and maximum information entropy slices as the input

of MRI images; at the same time, clinical information modality is introduced to obtain better classification performance.

- A comprehensive evaluation is conducted on ADNI dataset. Experiments show that our method achieves best classification performance in terms of accuracy, compared with most state-of-the-art methods.

The rest of the paper is structured in sections and represented as follows. In Section II, related work describes the research status of Alzheimer’s disease classification in detail. Section III introduces the structure and algorithm of the AMIM model. Section IV introduces the classification performance of the model on the ADNI dataset. Section V discusses the performance analysis of different views. Section VI concludes the paper.

## II. RELATED WORK

With the rapid development of deep learning since 2012, there are more and more researches on the diagnosis of Alzheimer’s disease. Researches based on Alzheimer’s disease classification can be divided into the following branches according to input: ROI level, Patch level, 3D Subject level and 2D slice level. With ROI models [22], [23], manual selection of regions is required to extract the region of interest of the original brain image as the input of CNN model, which is a time-consuming task. With patch models [19]–[21], multiple patches can be obtained from the entire 3D MRI, but there is a problem of data overlap. It is much more straightforward and desirable to use the entire image as input. At the 3D subject level, Korolev et al. [17] adopted 3D VGG and 3D ResNet as the backbone network for feature extraction, but the classification accuracy was only more than 80. Spasov et al. [18] proposed a method combining 3D MRI with clinical information, which can obtain good classification results. However, regardless of single mode or multi-mode 3D MRI, there is a large amount of calculation and long running time. In 2D slice classification method, it can reduce the number of hyperparameters to a certain extent. Due to the small sample size of medical dataset, Hon et al. [11] proposed to apply transfer learning to the classification of Alzheimer’s disease. 32 slices were taken from each object as the dataset, and the model performed well. However, this result was only for the image level, without considering the subject level. Islam et al. [12] proposed a deep convolutional neural network for the diagnosis of Alzheimer’s disease using brain MRI data analysis, and obtained good classification results. Zhang and colleagues [13] performed a systematic evaluation of CNN models with different structures and capacities, and the experimental results showed that the advanced structural models with medium capacity performed better than the models with maximum capacity. Good results have also been obtained. However, these methods are based on 2D images and cannot contain all the information of brain scan. They ignore the spatial information of 3D. The same situation exists in other slice classification studies [14]–[16]. According to the above analysis, we propose a new AD classification network architecture, AMIM, combining 2D-3D MRI and clinical information.

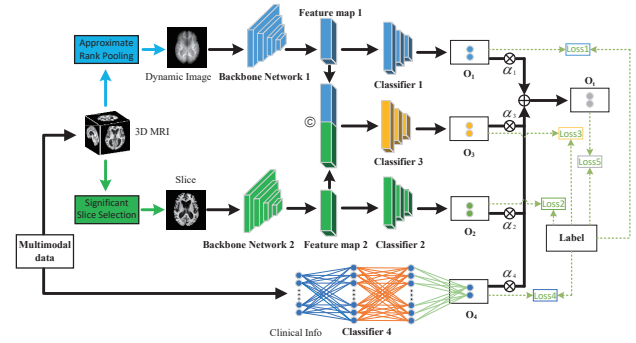


Fig. 1. The AMIM for alzheimer’s disease classification based on MRI and clinical information

## III. METHODS

In this paper, we propose an AMIM model that combines 2D-3D MRI to solve the problem of missing 3D spatial information for slice based classification. Clinical information is also introduced as the input of another modality. The model uses an adaptive weighted method to learn the weight shares of different classifier. Its architecture is shown in Fig. 1. Our proposed method is flexible and can in principle integrate other imaging modalities, such as positron emission computed tomography (PET), as well as other different clinical datasets. With the inspiration of the idea of transfer learning, we use the classical neural network pre-trained by ImageNet and removing the last classification layer as the backbone network of feature extraction [24]. Resnet18 as a backbone network will be introduced here.

In the following, we present our method in four parts. First, we introduce the dynamic image generation method and the maximum information entropy slicing method, respectively. Then, we present the adaptive weighted multimodal integration method. Finally, we introduce the training and optimization.

### A. Dynamic Image

In the non-medical field, a popular method to represent a series of images is to apply a temporal pooling operator to the features extracted at individual images, for instance, temporal templates [25], ranking functions [26] and other traditional pooling operator [27]. We use the Z-dimension of the 3D MRI as the temporal dimension of the video to extract a fixed slice representation of each object. Since the extracted fixed representation retains all the dynamic characteristics of slices (i.e. the changes from slice to slice), we call it dynamic image. We calculate the coefficient  $\theta_t$  of slice  $I_t$  and assume that the feature vector of this slice is  $V_t$ . Multiply this coefficient with the average of all feature vectors from  $V_1$  to  $V_t$  to get the new feature vector and finally accumulate the new feature vectors to get the final dynamic image. See Fig. 2 for an example. The calculation formula is as follows.

$$I_D = \sum_{t=1}^T \theta_t \psi(V_t) \quad (1)$$

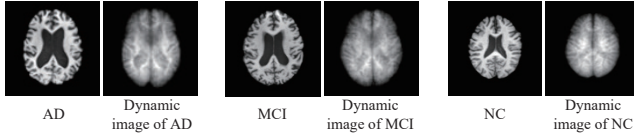


Fig. 2. MRI sample slices and corresponding dynamic images of AD, MCI and NC participants

where  $\psi(V_t) = \frac{1}{t} \sum_{\tau=1}^t V_\tau$  is the slices average of features up to slice  $t$ , and  $\theta_t = 2t - T - 1$  is the coefficient associated to feature  $\psi(V_t)$ .

The algorithm for processing dynamic images is shown in Algorithm 1.

---

**Algorithm 1** Algorithm for Obtaining Dynamic Images

---

**Input:** Origin 3D MRI:  $\{X_{MRI}^1, X_{MRI}^2, \dots, X_{MRI}^N\}$ .

**Output:** Dynamic images  $\{I_D^1, I_D^2, \dots, I_D^N\}$ .

- 1: **for**  $i = 1$  to  $N$  **do**
  - 2:  $I_O^j \leftarrow \text{slice}(X_{MRI}^i), j = \{1, 2, \dots, T\}$  // Obtaining original axial slices.
  - 3:  $V_j \leftarrow \text{quantify}(I_O^j)$  // Obtaining feature vectors.
  - 4: **for**  $t = 1$  to  $T$  **do**
  - 5:  $\theta_t \leftarrow 2t - T - 1$  // Calculating the coefficient  $\theta_t$  of slice  $I_O^t$ .
  - 6:  $\psi(V_t) \leftarrow \frac{1}{t} \sum_{\tau=1}^t V_\tau$  // Calculating the vector mean  $\psi(V_t)$  of the features vectors  $V_1$  to  $V_t$ .
  - 7: **end for**
  - 8:  $I_D^i \leftarrow \sum_{t=1}^T \theta_t \psi(V_t)$  // All new feature vectors are summed to get the final dynamic image by Eq.(1).
  - 9: **end for**
- 

### B. Slice with Maximum Information

Typically, there are a large number of slices to choose from in 3D MRI scan. One method of slices selection is to manually select slices based on the highest similarity of anatomical features without knowing the clinical diagnosis information [28]. However, this approach needs to be chosen by professionals, which will cost a lot of labor and be subjective. Instead, we use image entropy to extract the most informative slices to train the network. Therefore, we will calculate the image entropy of each slice. Generally speaking, for a set of  $M$  symbols with probabilities  $P_1, P_2, \dots, P_m$ . Entropy can be calculated as [29]:

$$H = - \sum_{i=0}^{255} P_i \log P_i \quad (2)$$

where  $H$  is the one-dimensional gray-scale entropy and  $P_i$  is the proportion of grayscale value  $i$ .

### C. Adaptive Weighted Multimodal Integration

In this section, we introduce the composition of classifiers and the adaptive weighted integration of the classifier, respectively.

For these classifiers we use the same composition structure. We add relu activation function and dropout between each layer of mapping to reduce the potential overfitting risks. Specifically the feature images are first dropout, and then through three layers of mapping, relu and dropout are added between each layer of mapping. Finally, softmax activation function is used to obtain the category probability value. Dropout set to 0.5. The output expression of each of these classifiers is shown as follows:

$$O_i = \varphi^i(F_i, C_i) \quad i \in \{1, 2, 3\} \quad (3)$$

where  $F_i$  stands for the input of the feature map,  $C_i$  is the weight parameter of the  $i$ -th classifier,  $\varphi^i(F_i, C_i)$  represents a function to be learned in an effort to transform the input,  $F_i$ , to probability value  $O_i$ .

$$O_4 = \varphi^4(X_{cli}, C_4) \quad (4)$$

where  $O_4$  is the probability value in the 4-th classifier,  $X_{cli}$  is the clinical information.  $C_4$  is the weight parameter of the 4-th classifier.  $\varphi^4$  denotes the operation of the 4-th classifier. We normalized for clinical characteristics, i.e. demographic, neuropsychological, and the apolipoprotein E (APOE4) genotyping data. They all followed the same feature scaling procedure, with values normalized between [0, 1] for each independent clinical factor.

We propose an ensemble learning method with adaptive weights to improve the performance of the model and the confidence of prediction. For the probability values of multiple classifiers, we use the integration method of soft voting for the final output. Suppose we have  $M$  classifiers, the soft voting can be computed as:

$$O_t = \sum_{i=1}^M \alpha_i O_i \quad (5)$$

where  $O_i$  is the probability value in the  $i$ -th classifier,  $\alpha_i$  represents the weight given to the  $i$ -th classifier.  $O_t$  stands for the total output after overall soft voting integration. First we initialize the weights. To automatically compute the hyperparameter  $\alpha_i$ , we use a simple but effective approach: setting the hyperparameter  $\alpha_i$  as a trainable parameter in order to automatically and adaptively coordinate the importance learning of each attribute task. When multiple branch tasks are learned simultaneously, the ‘‘important’’ branches should be given high weights (i.e.  $\alpha_i$ ) to increase the loss size of the corresponding branch. We take a small learning rate for updating the network parameters and automatically learn the score weights for different classifiers.

### D. Training and Optimization

We use cross-entropy as the loss function. We construct loss function for the output of individual classifier. The loss function is:

$$L_i = - \frac{1}{N} \sum_{j=1}^N \left[ y_j \cdot \log(O_i^j) + (1 - y_j) \cdot \log(1 - O_i^j) \right] \quad i \in \{1, 2, 3, 4\} \quad (6)$$

where the label  $y_j = 0$  indicates that sample  $j$  is a negative sample and  $y_j = 1$  indicates that sample  $j$  is a positive sample.  $N$  is the total number of samples in the data set.  $O_t^j$  denotes the output probability value of sample  $j$  of the  $i$ -th classifier. In the training phase, with the loss function constructed from the output of each classifier, i.e., Equation 6, we can optimize the network parameters in each of them.

For the backbone network, the weights are fixed in the first stage and are not optimized. The weights are unfrozen in the second stage, we first optimize the backbone network 1 and backbone network 2 under the loss functions constructed from the classifier 1 output and classifier 2 output, respectively. Then the backbone network is further optimized by the loss function 3.

After we get the output of each classifier, we can get the final output, which is Equation 5. Then the loss function of the integrated output is shown below:

$$L_5 = -\frac{1}{N} \sum_{j=1}^N \left[ y_j \cdot \log \left( O_t^j \right) + (1 - y_j) \cdot \log \left( 1 - O_t^j \right) \right] \quad (7)$$

where  $\alpha_i$  is the weight of the output probability values of the  $i$ -th classifier.  $O_t^j$  represents the output probability value of the soft voting integration of sample  $j$  through multiple classifiers. The hyperparameter  $\alpha_i$  is key part of the network. If we use the grid search method to obtain  $\alpha_i$ , this will be time consuming. We use the adaptive weighted method to update the network weights. The loss combines these weights as an integrate one to supervise the process of network training by adopting the back-propagation algorithm.

#### IV. EXPERIMENTS

##### A. Dataset

We use the publicly available dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) for our work. Specifically, we trained CNNs with the data from the "spatially normalized, masked, and N3-corrected T1 images" category. The brain MRI image size is  $110 \times 110 \times 110$ . Since a subject may have multiple MRI scans in the database, we use the most recent scan of each subject to avoid data leakage. All the data we used are summarized in Table I. Among them are 132 men and 92 women, aged between 55 and 90.3 years old. Friedman's ANOVA was used to test the difference in median age between groups, and the Fisher's exact test was used to test the gender interaction of the x group. These interactions are not statistically significant ( $p > .05$ ). And the following data: demographic data (age, gender, education level), neuropsychological cognitive assessment tests such as the Dementia Rating Scale (CDRSB), Alzheimer's Disease Assessment Scale (ADAS11, ADAS13), Rey Auditory Verbal Learning Test (RAVLT), as well as APOE4 genotyping. All data used in this study is from baseline assessments.

##### B. Evaluation Metrics

The proposed AMIM method mainly validates the AD classification (AD vs. NC), MCI classification (MCI vs. NC). The performance was evaluated using three metrics. Namely, accuracy, the percentage of correctly predicted samples; F1,

TABLE I. DEMOGRAPHICS OF THE ADNI DATASET

	F/M	Education	Age	APOE4	MMSE	CDRSB	ADAS11	ADAS13	RAVLT IR	RAVLT PF
CN	28/28	16.45±2.95	73.11±5.64	0.16±0.23	28.59±1.82	0.44±1.09	6.32±4.23	9.85±6.19	41.68±10.46	47.48±32.84
MCI	48/75	15.80±2.67	72.95±7.39	0.34±0.34	26.29±3.34	2.23±1.98	11.25±6.45	17.64±9.37	31.50±11.58	66.98±33.84
AD	16/29	15.37±2.89	74.66±8.54	0.39±0.33	21.67±3.94	5.81±2.40	24.07±7.48	34.78±9.01	19.56±6.71	93.68±13.35

the harmonic mean of precision (Eq. 9) and recall (Eq. 10); Area Under Curve (receiver operating characteristic curve determined by true positive rate and false positive rate). These metrics are defined as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

where TP, TN, FP and FN stands for true positive, true negative, false positive and false negative, respectively. A higher value indicates better performance.

In the following, experiments are conducted to evaluate the performance of the proposed method. Specifically, Section IV-D focuses on testing the impact of unimodal and multimodal on the performance metrics of the experiments, respectively. Section IV-E1 aims to analyze the classification performance of the same dataset in different methods.

##### C. Complements

We use five-fold cross-validation for experiments. Since the proportion of data samples is unbalanced, a weighted loss function is used to ensure the balance of the samples. The loss function uses a cross-entropy loss function. Using the Adam optimizer, the learning rate is 10 to the negative 5-th power, except that the learning rate of  $L_5$  is adjusted to  $5 \times 10^{-6}$ . The classifier performs linear mapping, it will perform a dropout of 0.5 to prevent overfitting. To activate relu, the last layer uses the softmax function to output the probability value of the category.

##### D. Single Modality vs. Multiple Modalities

In this section, the effects of unimodal and multimodal data on the model are presented separately. For MRI, resnet18, which has been pre-trained and removed the last classification layer, is used as the backbone network to extract features. The output of clinical information is obtained through a multi-layer perceptron network model. We performed single-modal experiments on dynamic images, slices, and clinical information. As shown in the Table II, we can see that in the column of AD versus NC, the evaluation index of clinical information is very high. Thus we made t-SNE visualization for the data, as shown in Fig. 3. The data distribution of AD group and NC group is shown on the left. It is found that the two types of data have obvious dividing lines. The distributions of the MCI and NC groups on the right do not have obvious dividing



TABLE II. COMPARISON OF PERFORMANCE OF SINGLE-MODAL AND MULTIMODAL CLASSIFICATION METHODS

Methods	AD vs. NC (%)			MCI vs. NC (%)		
	ACC	AUC	F1	ACC	AUC	F1
Dynamic	90.00	91.90	88.63	77.10	74.45	84.61
Slice	83.14	82.59	83.09	71.00	62.34	80.76
Clinical Info	97.00	98.03	97.41	77.67	78.65	85.53
Combined	92.00	93.62	91.06	74.29	72.39	82.26
Fusion	<b>99.00</b>	<b>99.12</b>	<b>98.67</b>	<b>81.63</b>	<b>82.41</b>	<b>87.06</b>

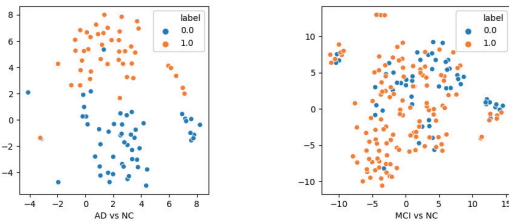


Fig. 3. Illustration of t-SNE result for cognitive scores. Blue dots: NC group. Orange dots: AD/MCI group

lines, and their evaluation indexes in Table II are not very satisfactory. This remains consistent with our results. Then we take the dynamic images and slices of medical imaging as input to obtain the integrated results of medical imaging. Specifically, the decision fusion of Output 1, Output 2 and Output 3 of the model in Fig. 1 is carried out. The specific experiment is introduced in the subsequent result chapter. It can be seen from the table that the integration results of slices combined with pictures of 3D spatial information changes are improved compared with single-mode medical imaging. Finally, we proposed the method, in which AD/NC accuracy reached 99.00%, MCI/NC accuracy also reached 81.63%.

### E. Comparison of Different Methods

1) *Comparison With baseline methods:* We compared several other methods on the same dataset. Korolev et al. proposed a deep three-dimensional convolutional neural network structure for brain MRI scan classification [17]. In this work [18], structural magnetic resonance imaging (MRI), demography, neuropsychology and APOE4 gene were used as data inputs, and 3D separable convolutional layers were used as backbone networks for classification. Xing et al converted the 3D full image into 2D dynamic image, and then took the classical neural network and attention mechanism as the network model [30]. For these baseline methods, we maintain the parameter settings of the original paper. The results of evaluation indicators are shown in Table III. It can be found that our proposed method is the best in most indicators.

2) *Comparison with state-of-the-art methods:* In this section, we focus on comparing the classification performance of other widely used methods. The work investigated [31]–[37] and other MRI monomodal classification performance, see Table IV below. Multi-modality [32]–[34], [36], [38]–[41] includes MRI + PET, MRI + PET + biomarkers, MRI + DTI, and MRI + Cognitive scores, as shown in Table V.

TABLE III. COMPARISON OF PERFORMANCE OF DIFFERENT CLASSIFICATION METHODS

Methods	Type	AD vs. NC (%)			MCI vs. NC (%)		
		ACC	AUC	F1	ACC	AUC	F1
Korolev et al. [17]	3D	70.00	69.54	67.57	73.75	67.40	82.64
Spasov et al. [18]	3D+Cli Info	98.00	<b>99.38</b>	97.70	80.62	<b>83.91</b>	84.52
Xing et al. [30]	3D	86.25	85.93	86.08	78.13	70.25	84.58
Proposed	3D+2D+Cli Info	<b>99.00</b>	99.12	<b>98.67</b>	<b>81.63</b>	82.41	<b>87.06</b>

TABLE IV. PERFORMANCE COMPARISON OF DIFFERENT METHODS IN SINGLE-MODALITY STUDY

Study	Modality	Group			Method	Accuracy(%)	
		AD	MCI	NC		AD vs. NC	MCI vs. NC
Liu et al. [31]	MRI	97	234	128	multi-view: multi-template:SVM	93.83	-
Zhu et al. [32]	MRI	51	99	52	ROI-based:Relational regularization feature selection	93.70	79.70
Liu et al. [33]	MRI	93	204	100	Patch-based: Cascaded CNNs	84.97	-
Aderghal et al. [34]	MRI	188	399	228	Hippocampus:CNN; Transfer learning	90.00	72.50
Basaia et al. [35]	MRI	295	510 <sup>a</sup> + 253 <sup>b</sup>	352	Patch-based:CNN	99.20	76.10 <sup>c</sup> 87.10 <sup>d</sup>
Shao et al. [36]	MRI	160	460	160	Hypergraph; Multikernel SVM	88.30	69.14
Al-Khuzai et al. [37]	MRI	170	-	70	Slices-based:CNN	99.30	-

Group:  $a^a + b^b$ : the number of sMCI is  $a$  and the number of pMCI is  $b$ ; Accuracy:  $c^c$ : the accuracy of sMCI/NC is  $c$ ;  $d^d$ : the accuracy of pMCI/NC is  $d$ .

TABLE V. PERFORMANCE COMPARISON OF DIFFERENT METHODS IN MULTI-MODALITY STUDY

Study	Modality	Group			Method	Accuracy(%)	
		AD	MCI	NC		AD vs. NC	MCI vs. NC
Zhang et al. [38]	MRI+PET+CSF	51	99	52	ROI-based: Kernel combination	93.20	76.40
Zhu et al. [32]	MRI+PET	51	99	52	ROI-based:Relational regularization feature selection	95.70	79.90
Shi et al. [39]	MRI+PET	51	99	52	ROI-based: Multimodal SDPN	97.13	<b>87.24</b>
Liu et al. [33]	MRI+PET	93	204	100	Patch-based: Cascaded CNNs	93.26	74.34
Aderghal et al. [34]	MRI+DTI	48	108	58	Hippocampus; CNN; Transfer learning	92.50	80.00
Shi et al. [40]	MRI+PET+CSF	51	99	52	ROI-based: Coupled boosting; Coupled metric ensemble	94.85	79.88
Shao et al. [36]	MRI+PET	160	460	160	Hypergraph; Multi-kernel SVM	92.51	82.53
Hett et al. [41]	MRI+Cognitive scores	130	216	213	Patch-based: Multiscale graphs	91.60	-
Proposed	MRI+Cognitive scores	45	123	56	CNN:Transfer learning; Ensemble	<b>99.00</b>	81.63

CSF = Cerebrospinal fluid

For AD/NC classification, the accuracy of the single-modal methods listed in Table IV has classification results below 90.00%, while the accuracy of most multi-modal methods is above 90.00%. For MCI/NC classification, the accuracy of most single-modal methods is below 80.00%, while the accuracy of most multi-modal methods is above 80%. Among the listed studies, Zhu et al. [32], Liu et al. [33], Aderghal et al. [34] and Shao et al. [36] performed single-modal and multi-modal tests on the proposed method. The results show that, compared with single-modal data, the use of multi-modal data can obtain higher classification accuracy. In addition, our method achieves the best performance of 99.00% in the AD/NC classification and 81.63% in the MCI/NC classification with resnet18 as the backbone network. It is worth noting that, due to potential differences in data selection, preprocessing and even data set division, the results obtained by different methods are actually incomparable. The purpose of the comparison is only to provide an overview of other results and to show the baseline of existing methods.

3) *Different backbone neural networks:* We performed other backbone neural network training. To evaluate the classification performance of different backbone neural networks, we used a five-fold cross-validation strategy to calculate the classification performance. Specifically, the entire subject sample set was divided into five subsets equally, and the subject samples within one subset were selected as test samples each time, and all remaining subject samples within the other four subsets were used to train the classifiers. This process was repeated five times independently to avoid any bias introduced by randomly dividing the dataset in cross-validation. We take the average of the three experiments as the final result of the data. The results are shown in the Table VI and Table VII.

TABLE VI. PERFORMANCE OF DIFFERENT BACKBONE NEURAL NETWORKS FOR AD VS. NC (%)

Methods	ACC	AUC	F1	PRECISION	RECALL	AP
AMIM+resnet18	98.33	98.75	98.01	96.85	99.48	96.67
AMIM+resnet34	98.33	98.75	98.03	97.50	98.79	96.95
AMIM+resnet50	98.67	99.27	98.35	97.50	99.39	97.23
AMIM+resnet101	99.00	98.68	98.67	97.50	1.0	97.50
AMIM+vgg11	98.33	99.12	98.08	97.50	98.88	97.05
AMIM+googlenet	97.00	99.19	96.67	97.38	96.20	95.72
AMIM+alexnet	98.00	98.75	97.38	96.09	99.05	95.60

TABLE VII. PERFORMANCE OF DIFFERENT BACKBONE NEURAL NETWORKS FOR MCI VS. NC (%)

Methods	ACC	AUC	F1	PRECISION	RECALL	AP
AMIM+resnet18	80.51	80.73	86.72	84.52	95.42	81.86
AMIM+resnet34	82.52	82.78	88.18	86.51	96.85	83.65
AMIM+resnet50	78.61	79.98	85.30	87.18	95.76	83.10
AMIM+resnet101	78.40	80.33	85.54	86.52	96.15	81.66
AMIM+vgg11	80.09	80.74	86.97	82.44	97.95	80.83
AMIM+googlenet	79.71	77.80	86.24	84.02	97.11	81.45
AMIM+alexnet	82.86	82.71	88.19	85.57	93.90	83.84

## V. DISCUSSION

### A. Performance analysis of different views

3D MRI exists three views (axial, coronal and sagittal view). We used the axial view. The different views of the 3D global information and the maximum information slice are shown in Fig. 4. Pan et al. [42] proposed a Multi-view Separable Pyramid Network (MiSePyNet), in which representations are learned from axial, coronal and sagittal views of PET scans so as to offer complementary information and then combined to make a decision jointly. The experimental results show that the performance of the axial view is the best and multi-view fusion effect is better than the single-view. Next, this paper will discuss and analysis the classification performance of different views through experiments.

We performed experiments for different views by using the same parameter settings. The overall experimental results of AD versus NC and MCI versus NC are shown in Table VIII and Table IX. The classification performance of different views for each model has been given in the table. The best evaluation metrics are shown in bold. The experimental results show that AD versus NC and MCI versus NC experimental

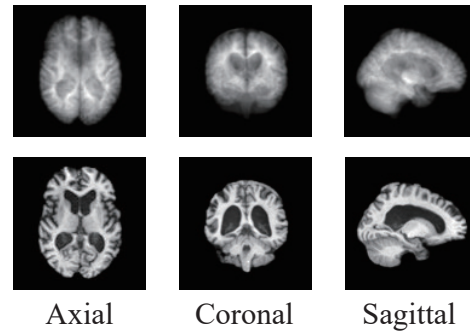


Fig. 4. Different views of the global information image and the maximum information slice

TABLE VIII. PERFORMANCE OF DIFFERENT VIEWS FOR AD VS. NC (%)

Type	Views	ACC	AUC	F1
Slice	Axial view	83.14	82.59	<b>83.09</b>
	Coronal view	<b>85.24</b>	<b>87.49</b>	82.39
	Sagittal view	81.19	86.23	79.41
Dynamic	Axial view	<b>90.00</b>	91.90	<b>88.63</b>
	Coronal view	86.05	87.49	86.44
	Sagittal view	89.05	<b>93.82</b>	88.58
Fusion	Axial view	<b>92.00</b>	<b>93.62</b>	<b>91.06</b>
	Coronal view	86.10	84.33	85.45
	Sagittal view	86.14	85.48	85.39

TABLE IX. PERFORMANCE OF DIFFERENT VIEWS FOR MCI VS. NC (%)

Type	Views	ACC	AUC	F1
Slice	Axial view	71.00	62.34	80.76
	Coronal view	<b>72.10</b>	<b>67.71</b>	<b>81.35</b>
	Sagittal view	69.89	63.97	80.89
Dynamic	Axial view	<b>77.10</b>	<b>74.45</b>	<b>84.61</b>
	Coronal view	72.65	66.63	82.23
	Sagittal view	72.10	62.72	82.66
Fusion	Axial view	<b>74.29</b>	<b>72.39</b>	<b>82.26</b>
	Coronal view	71.54	55.46	81.14
	Sagittal view	70.44	54.81	79.65

classification performance are similar. In the model with only slices as input, the overall evaluation metrics of coronal view are the best, the evaluation metrics of the other two views have small differences with them. In the model with only dynamic images as input, the AUC and F1 metrics of the axial view are the highest. In the whole hybrid model of dynamic image and slices, all evaluation metrics of the axial plane is the highest.

For the overall results obtained from these three different views, we can find that the classification performance of each view is not very different and the axial view is relatively better. In this paper, we only conducted experiments for one view. In the future, we will analyze three views of the image together. Different views show different information and how to obtain more comprehensive information plays a more important role

for Alzheimer's disease diagnostic.

## B. Data Pairing

The dataset we used is from the "Spatially normalized, masked and N3-corrected T1 images" category in the ADNI public dataset. The dataset in this category contain paired MRI and clinical information. However, for the other categories of dataset in the ADNI, there were cases where subjects had not undergone the MMSE, ADAS11 examination. For patients with clinical information, we effectively combined the clinical information, which helped to improve the classification performance of the model. In the future, we will try to improve the diagnostic performance of ADNI with only some of the subjects' basic information (gender, age, etc.).

## VI. CONCLUSION

In this paper, we proposed a multimodal adaptive weighted model, which takes global information images, maximum information slices and clinical information as multimodal inputs for the first time. Our model can effectively solve the problem of missing global information in slice classification. At the same time, the use of image information entropy selection slices can solve the subjective uncertainty of human selection. Using an adaptive weighting method to optimize the weights, it can combine the weights of different models more accurately than the grid search method. Our model achieves the best results in terms of classification performance, compared with the latest methods. The combination of medical images and clinical information for Alzheimer's disease classification is the future trend. Next, we will try to investigate how to better combine clinical information with medical images.

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their help. This work was supported by the National Natural Science Foundation of China (Grant No. 62076044) ; the Key Project of Chongqing Natural Science Foundation in China (Grant No. cstc2019jcyjzdxmX0011); Chongqing Graduate Research Innovation Project of China (Grant No. CYS21307).

## REFERENCES

- [1] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, "Forecasting the global burden of Alzheimer's disease," *Alzheimer's & Dementia*, vol. 3, no. 3, pp. 186–191, 2007.
- [2] W. Jagust, "Vulnerable neural systems and the borderland of brain aging and neurodegeneration," *Neuron*, vol. 77, no. 2, pp. 219–234, 2013.
- [3] J. Cummings, G. Lee, A. Ritter, M. Sabbagh, and K. Zhong, "Alzheimer's disease drug development pipeline: 2019," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 5, no. 1, pp. 272–293, 2019.
- [4] A. Atri, "Current and future treatments in Alzheimer's disease," in *Seminars in neurology*, vol. 39, no. 02, 2019, pp. 227–240.
- [5] M. A. Ebrahimighahnavieh, S. Luo, and R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review," *Computer Methods and Programs in Biomedicine*, vol. 187, p. 105242, 2020.
- [6] F. Li, L. Tran, K.-H. Thung, S. Ji, D. Shen, and J. Li, "A Robust Deep Model for Improved Classification of AD/MCI Patients," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 5, pp. 1610–1616, 2015.
- [7] H.-I. Suk, C.-Y. Wee, S.-W. Lee, and D. Shen, "State-space model with deep learning for functional dynamics estimation in resting-state fmri," *NeuroImage*, vol. 129, pp. 292–307, 2016.
- [8] Y. Li, F. Meng, and J. Shi, "Learning using privileged information improves neuroimaging-based cad of alzheimer's disease: a comparative study," *Medical & Biological Engineering & Computing*, vol. 57, pp. 1605–1616, 2019.
- [9] M. I. Razzak, S. Naz, and A. Zaib, *Deep Learning for Medical Image Processing: Overview, Challenges and the Future*, N. Dey, A. S. Ashour, and S. Borra, Eds. Cham: Springer International Publishing, 2018.
- [10] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [11] M. Hon and N. Khan, "Towards Alzheimer's disease classification through transfer learning," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Los Alamitos, CA, USA: IEEE Computer Society, 2017, pp. 1166–1169.
- [12] J. Islam and Y. Zhang, "Early Diagnosis of Alzheimer's Disease: A Neuroimaging Study with Deep Learning Architectures," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 1962–19622.
- [13] F. Zhang, B. Pan, P. Shao, P. Liu, S. Shen, P. Yao, and R. X. Xu, "A Single Model Deep Learning Approach for Alzheimer's Disease Diagnosis," *Neuroscience*, vol. 491, pp. 200–214, 2022.
- [14] S. Luo, X. Li, and J. Li, "Automatic Alzheimer's Disease Recognition from MRI Data Using Deep Learning Method," *Journal of Applied Mathematics and Physics*, vol. 05, pp. 1892–1898, 2017.
- [15] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images," *Cognitive Systems Research*, vol. 57, pp. 147–159, 2019.
- [16] J. Liu, M. Li, Y. Luo, S. Yang, W. Li, and Y. Bi, "Alzheimer's disease detection using depthwise separable convolutional neural networks," *Computer Methods and Programs in Biomedicine*, vol. 203, p. 106032, 2021.
- [17] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 835–838.
- [18] S. Spasov, L. Passamonti, A. Duggento, P. Liò, and N. Toschi, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease," *NeuroImage*, vol. 189, pp. 276–287, 2019.
- [19] D. Cheng and M. Liu, "Cnns based multi-modality classification for AD diagnosis," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2017, pp. 1–5.
- [20] M. Liu, D. Cheng, K. Wang, and Y. Wang, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, no. 3, pp. 295–308, 2018.
- [21] S. Qiu, P. S. Joshi, M. I. Miller, C. Xue, X. Zhou, C. Karjadi, G. H. Chang, A. S. Joshi, B. Dwyer, S. Zhu, M. Kaku, Y. Zhou, Y. J. Alderazi, A. Swaminathan, S. Kedar, M.-H. Saint-Hilaire, S. H. Auerbach, J. Yuan, E. A. Sartor, R. Au, and V. B. Kolachalama, "Development and validation of an interpretable deep learning framework for Alzheimer's disease classification," *Brain*, vol. 143, no. 6, pp. 1920–1933, 2020.
- [22] J. M. Rondina, L. K. Ferreira, F. L. de Souza Duran, R. Kubo, C. R. Ono, C. C. Leite, J. Smid, R. Nitrini, C. A. Buchpiguel, and G. F. Busatto, "Selecting the most relevant brain regions to discriminate Alzheimer's disease patients from healthy controls using multiple kernel learning: A comparison across functional and structural imaging modalities and atlases," *NeuroImage: Clinical*, vol. 17, pp. 628–641, 2018.
- [23] B. Duraisamy, J. V. Shanmugam, and J. Annamalai, "Alzheimer disease detection from structural MR images using FCM based weighted probabilistic neural network," *Brain imaging and behavior*, vol. 13, no. 1, pp. 87–110, 2019.
- [24] S. Kornblith, J. Shlens, and Q. V. Le, "Do better imagenet models transfer better?" in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2656–2666.

- [25] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257–267, 2001.
- [26] B. Fernando, E. Gavves, M. José Oramas, A. Ghodrati, and T. Tuytelaars, "Modeling video evolution for action recognition," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5378–5387.
- [27] H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould, "Dynamic Image Networks for Action Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3034–3042.
- [28] S. Qiu, G. Chang, M. Panagia, D. Gopal, R. Au, and V. Kolachalama, "Fusion of deep learning models of MRI scans, Mini-Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 10, 2018.
- [29] C. Studholme, D. Hill, and D. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition*, vol. 32, no. 1, pp. 71–86, 1999.
- [30] X. Xing, G. Liang, H. Blanton, M. U. Rafique, C. Wang, A.-L. Lin, and N. Jacobs, "Dynamic Image for 3D MRI Image Alzheimer's Disease Classification," in *Computer Vision – ECCV 2020 Workshops*, A. Bartoli and A. Fusiello, Eds. Cham: Springer International Publishing, 2020, pp. 355–364.
- [31] M. Liu, D. Zhang, E. Adeli, and D. Shen, "Inherent Structure-Based Multiview Learning With Multitemplate Feature Representation for Alzheimer's Disease Diagnosis," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1473–1482, 2016.
- [32] X. Zhu, H.-I. Suk, L. Wang, S.-W. Lee, and D. Shen, "A novel relational regularization feature selection method for joint regression and classification in AD diagnosis," *Medical Image Analysis*, vol. 38, pp. 205–214, 2017.
- [33] M. Liu, D. Cheng, K. Wang, and Y. Wang, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, no. 3, pp. 295–308, 2018.
- [34] K. Aderghal, A. Khvostikov, A. Krylov, J. Benois-Pineau, K. Afdel, and G. Catheline, "Classification of Alzheimer Disease on Imaging Modalities with Deep CNNs Using Cross-Modal Transfer Learning," in *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, 2018, pp. 345–350.
- [35] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, and M. Filippi, "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage: Clinical*, vol. 21, p. 101645, 2019.
- [36] W. Shao, Y. Peng, C. Zu, M. Wang, and D. Zhang, "Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease," *Computerized Medical Imaging and Graphics*, vol. 80, p. 101663, 2020.
- [37] F. E. Al-Khuzai, O. Bayat, and A. D. Duru, "Diagnosis of Alzheimer disease using 2D MRI slices by convolutional neural network," *Applied Bionics and Biomechanics*, vol. 2021, 2021.
- [38] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.
- [39] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 173–183, 2018.
- [40] Y. Shi, H.-I. Suk, Y. Gao, S.-W. Lee, and D. Shen, "Leveraging Coupled Interaction for Multimodal Alzheimer's Disease Diagnosis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 1, pp. 186–200, 2019.
- [41] K. Hett, V.-T. Ta, I. Oguz, J. V. Manjón, and P. Coupé, "Multi-scale graph-based grading for Alzheimer's disease prediction," *Medical Image Analysis*, vol. 67, p. 101850, 2021.
- [42] e. a. Pan X, Phan T L, "Multi-view separable pyramid network for ad prediction at mci stage by 18 f-fdg brain pet imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 81–92, 2020.